JOINT ATTENTION IN INFANCY: DEFINITIONS, SENSORY-MOTOR BASIS AND DEVELOPMENTAL CONSEQUENCES

Catalina Suarez Rivera

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Psychological and Brain Sciences,
Indiana University
July 2019

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Doctoral Committee

Linda B. Smith, Ph.D.

Chen Yu, Ph.D.

Daniel P. Kennedy, Ph.D.

John Kruschke, Ph.D.

May 31, 2019

Acknowledgements

I would like to thank my mentors, Dr. Linda B. Smith and Dr. Chen Yu for teaching me about science and also life, for helping me grow so much by dedicating their time and effort in every step of this beautiful process. Thanks Linda for instructing me to trust the process, for inspiring me to love writing and for being an extraordinary, strong and caring mentor.

I would also like to thank my parents Pedro and Elizabeth, for their love and support, for asking me about my work and for celebrating both big accomplishments and everyday small victories. Thank you for inspiring by example to do my best work. Thanks for raising me to accept challenges and persist despite difficulties. My parents and my grandparents taught me to value education and enjoy learning about the world with curiosity. I am forever grateful.

I am forever grateful with my brother Nicolas who patiently listens and wisely advises me. I am so grateful I count on him and I pray we continue to support each other in life.

All past and present members of the Cognitive Development lab and the Computational Cognition and Learning lab made the journey so pleasant through lab meetings, socials and soccer games that I cherish. It has been an honor to be part of such a collaborative and intellectually rich group. Thanks in particular to my best friend Dr. Lauren K. Slone for supporting me and helping me with everything. Thank you for being a role model.

I am grateful with the community of Saint Paul Catholic Center. In particular, Fr. Joseph Minuth and Fr. Raymond Bryce helped with the one thing that turned my life around, the source of love, hope, peace and faith. Thank for inspiring me to grow closer to God.

Finally, thanks to all faculty and research staff including Char Wozniak, Misty Bennett, Lana Fish, Patricia Crouch and JeanneMarie Hebb. Thanks for the support, for being kind and encouraging. Thank you for caring for graduate students so much, we truly appreciate it.

Catalina Suarez Rivera

JOINT ATTENTION IN INFANCY: DEFINITIONS, SENSORY-MOTOR BASIS AND DEVELOPMENTAL CONSEQUENCES

Successful social interactions require the coupling of the bodies and minds of the participants. *Joint visual attention* to the same target is a critical component of this coupling. By 9 months of age, as measured both in trial-based experiments and in continuous naturalistic play, infants engage in joint attention with an adult. Infants' early joint attention skills, measured in both contexts, are strong predictors of future developmental outcomes including language development. Thus, the consensus assumption is that shared skill sets support joint attention in trial-based experiments and in free play with a social partner. Growing evidence raises doubts about this assumption since the different contexts for joint attention appear to elicit the phenomenon differently. Research is needed to fully understand the phenomenon across contexts and determine whether individual differences in trial-based joint attention are predictive of individual differences in free play with the parent. Furthermore, research is needed to examine joint attention moment-to-moment during play in order to examine the real world infant and parent behaviors that occur during these moments of parent-infant coupling and can point to mechanisms that can potentially link joint attention to infant learning and development. The previous studies that show early joint attention predicts later language development show correlational evidence that the coupling of infants with mature social partners matters for development, and thus the specific pathways through which joint attention benefits infant learning and development are not known. The aims of the dissertation are 1) to understand how joint attention between infants and social partners is achieved in discrete trial-based and free play contexts, and whether individual differences in the contexts are related, and 2) to examine the developmental consequences of joint attention in the context of free play in order to begin building a pathway between joint attention and infant learning and development. Together, three studies show that coordinated visual attention occurring during continuous naturalistic parent-infant play supports the infant's own control of visual attention both in-the-moment and eight months later, but that measures of joint attention obtained during naturalistic play do not tap the same underlying processes measured in trial-based contexts for joint attention. I discuss implications of this work for the theoretical claims on joint attention and suggest a likely pathway through which joint attention achieved in the context of free play with the parent may support language learning as well as academic achievement and self-regulation. The results demonstrate one instance of the larger and general problem in science: when clean and controlled findings in the laboratory fail to translate to the complexity of real-world life.

Linda B. Smith, Ph.D.

Chen Yu, Ph.D.

Daniel P. Kennedy, Ph.D.

John Kruschke, Ph.D.

Table of Contents

Chapter 1

Introduction to the Dissertation

The quality of parent-infant social interactions is well known to have long-lasting effects on the development of children's social skills (Landry, Smith & Swank, 2006), self-regulation (Fay-Stammbach, Hawes, & Meredith, 2014), language (Tamis-LeMonda, Bornstein & Baumwell, 2001), and school achievement (Grolnick & Ryan, 1998). The quality of these social interactions, are, of course not solely the product of the parent's own behavior nor that of the child, but a dyadic property. One component of early social interactions that predicts future development is known as joint attention (Tomasello & Farrar, 1986; Yu, Suanda & Smith, 2019). Shared attention to a visual target is by definition a dyadic property but the original and dominating approach within researchers has been to focus on the skills that the infant needs in following gaze cues from a social partner. Most of the research on this phenomenon has been carried out in the context of a scripted interaction. These traditional measures of the infants' skill are obtained from discrete trial-based experiments taking place in the laboratory (e.g., Brooks & Meltzoff, 2005; Mundy, Delgado, Block, Venezia, Hogan & Scibert, 2003), a context that is ideal to provide fair estimates from all infants since all subjects get exposed to the same scripted interaction with the experimenter. The scripted interaction that is part of the research context defined for the experiment not only misses the dyadic component of joint attention but it also is unnatural.

Joint attention, two individuals look to the same object at the same time, also occurs in everyday contexts. A growing body of literature suggests that the phenomenon elicited in the context of discrete trial-based experiments is not representative of the phenomenon that occurs in everyday real-world contexts. Conversely, the paradigm of free parent-infant play session is representative of more natural contexts like the home and thus, provides insights into the properties of the moments of joint attention as they naturally occur between the parent and the infant. In the context of continuous free play sessions, infants follow social cues that contain manual contact with the attended object and engage in joint attention through "hand-following" pathways (Yu & Smith, 2013; 2017a). Moreover, infants play a dual role as initiators and followers of the joint attention moments (Yu & Smith, 2013). Finally, joint attention is defined as a *moment*, one that results from the fine sensory-motor coordination of the infant and parent within themselves, and between the two (Tasker & Schmidt, 2008; Smith & Yu, 2017a; 2017b). Joint attention elicited in discrete trial-based experiments looks strikingly different, as it does not involve holding cues and infants only follow with their eyes an experimenter's gaze that occurs in predefined trials with no relation to the infant's state (Brooks & Meltzoff, 2005; Mundy, Delgado, Block, Venezia, Hogan & Scibert, 2003). Joint attention in the context of discrete trial-based experiments is believed to be an infant ability that can be measured equally for all infants given the scripted interaction and standard behavior of the adult (Tasker & Schmidt, 2008). The evidence suggests that the phenomenon elicited in the context of trial-based experiments lacks, by design, the natural properties that describe how joint attention occurs in the real world.

Lab-based phenomena that are not representative of what occurs in the real world create serious problems for researchers in general (Rothwell, 2005) but especially for those interested in understanding how humans develop. Experiments help researchers reduce the complexity in the world, isolate variables to make causal inferences and propose mechanisms of development. But mechanisms are only valid for the study of development if they play out in the real world. Every day behavior is the driver of developmental change since effects seen in larger timescales of development are built from the accumulated changes happening in real time moment-to-moment (Smith & Thelen, 2003). Thus a crucial question is how clear, precise and predictive findings in laboratory relate to the real world phenomena we are trying to understand.

The paradigm of continuous free play session approximates more closely to the real world and it allows researchers to propose detailed pathways that connect *real world* behavior to *real world* infant learning and development. Recent research that uses the paradigm of continuous free play sessions, has shown that joint attention occurring in the context of every day play between the parent and the infant supports the maintenance of infant visual attention to the jointly attended object (Wass, Clackson, Georgieva, Brightman, Nutbrown & Leong, 2018a; Yu & Smith, 2016). The evidence suggests that joint attention supports infant sustained attention and because infant sustained attention in turn supports infant learning (Lansik & Richards, 1997; Richards & Casey, 1992; Ruff, 1986), Yu & Smith (2016) raised the hypothesis that infant sustained attention is part of the developmental pathway that connects joint attention to future learning and development. There are open questions that pertain to this hypothesis, in particular what are the parent behaviors that support infant sustained attention during moments of coordinated joint visual attention? And what is the role of coordinated joint attention moments for the development of self-controlled sustained attention? Research that addresses these questions by measuring joint attention in the context of continuous free play is needed to begin testing the exact mechanisms through which joint attention supports language learning and development. Previous studies that examine the role of joint attention for child development by using discrete trial-based measurements of the phenomenon, may have measured a phenomenon that is not related to the real world and have established a link between joint attention skills and later language ability that is purely correlational (Brooks & Meltzoff, 2005; Morales, Mundy &

Rojas, 1998; Mundy et al., 2007; Markus et al., 2000). As a result these previous studies cannot explain the exact ways in which joint attention may benefit *real world* infant learning and development.

The dissertation addresses important open questions in the field and has two specific aims: 1) to investigate the relation between measures of joint attention provided by standard discrete trial-based experiment and measures of joint attention provided by the context of free play in order to address the apparent *Joint Attention Disconnect*; 2) to examine the developmental consequences of infant-parent joint attention for *infant sustained attention* in order to begin building a path from joint attention to infant learning and development. As demonstrated by previous research, joint attention during everyday play supports infant sustained attention (Wass et al., 2018a; Yu & Smith, 2016) and in turn, sustained infant visual attention is associated with visual learning of the attended object and with future developmental outcomes including language ability, self-regulation and academic achievement (Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov, Pagani, Feinstein, Engel, Brooks-Gunn, Sexton, Duckworth & Japel, 2007; Kochanska, Murray, & Harlan, 2000; McClelland, Acock & Morrison, 2006; McClelland & Cameron, 2012; Welsh, Nix, Blair, Bierman & Nelson, 2010). Two thirds of this dissertation work focus on testing the hypothesis that coordinated joint visual attention during free play supports the development of infant's own control of attention.

For the purposes of clarity, joint attention elicited by the discrete trial-based experiment is referred to in this dissertation as **Joint Attention**. The term **Coordinated Joint Attention** is used in turn to refer to the phenomenon as it is elicited in the context of continuous free play sessions, while the term **joint attention** is used to refer to the abstract concept of the phenomenon without respect to the research context in which it is being measured by researchers.

The dissertation has three experimental chapters that are written as stand-alone papers formatted for publication. Chapter 2 investigates 9-month-olds Joint Attention and Coordinated Joint Attention in order to address whether there is a *Disconnect* between the two dominant approaches that are used to study joint attention. Are the individual differences captured by the discrete trial-based experiment predictive of infant behavior in the context of free play? This paper is not yet submitted for publication and presents a longer version of the to-be-submitted paper, taking the room allowable in a thesis to discuss the broader implications of the findings. The other two experimental chapters of the dissertation investigate *only* Coordinated Joint Attention because the context of free play, and not the context of discrete trial-based experiments, allows one to examine real world joint attention as well as its *real world* developmental consequences. Chapter 3 examines the multimodal nature of Coordinated Joint Attention and its real time effect on the infant's own ability to sustain visual attention. This paper is published in *Developmental Psychology.* Chapter 4 tests the long-term consequences of Coordinated Joint Attention that effectively extended infant attention in the moment. Together, Chapters 3-4 build the case for a mechanistic pathway that links social partners and infants' future development, a pathway described in the General Discussion of the dissertation along with other open questions and future work (Chapter 5).

Chapter 2

The Joint Attention Disconnect: Comparing babies in a discrete trial-based experiment and in a continuous free play session

The end-goal of many research programs is to establish *causal* links and pinpoint mechanisms with respect to the phenomenon of interest. Experiments that manipulate potential causes and control for extraneous factors are the gold standard for determining causes and specific mechanisms. Nonetheless, an increasingly large number of researchers are now carrying out their research in more naturalistic settings such as the home environment (e.g., Fausey, Jayaraman & Smith, 2016; Karasik, Tamis-LeMonda, Adolph & Bornstein, 2015; Roy, Frank, DeCamp, Miller & Roy, 2015; Tamis-LeMonda, Custode, Kuchirko, Escobar & Lo, 2018), where experimental control is reduced and necessary and sufficient causes are not determinable. In the same way, researchers are also leading initiatives to collect corpora that are home-based (VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, 2016; see Homeview project led by Dr. Linda Smith and PLAY led by Drs. Karen Adolph, Catherine Tamis-LeMonda and Rick Gilmore) and there are increasing discussions of in-principle limitations of basic science approaches for translation (Dahl, 2017; Kessler & Glasgow, 2011; Schmuckler, 2001; Rothwell, 2005; Tamis-LeMonda, Kuchirko, Luo, Escobar & Bornstein, 2017). One possible approach to this problem is both to try to understand –with precision—behaviors in real-world free-flowing multicausal contexts and to try to then link those behavioral measures to experimental findings. One possibility, however, is that much of the phenomena elicited in the laboratory links only weakly to the behaviors and internal processes that support behaviors in the world. Here we illustrate these larger points by unpacking joint attention in laboratory experiments and as it unfolds in free-flowing interactions, and showing that infant behavior elicited in the trial-based lab context is not related to behavior that occurs in the complexity of more naturalistic contexts. In conclusion, we argue that we need to embrace the complex causal pathways in the real world, and find ways to determine them, moving away from tightly controlled clean "petri dish" research designs.

The Translation Problem

Experiments are widely used *tools* that researchers use to uncover mechanisms and find causal relationships that link a cause with its effect. In a typical experiment, the phenomenon of interest is recreated, or specified, in a research context that is held constant across subjects and allows researchers to control –by turning on/off, or by varying degrees— the independent variable they hypothesize has a causal effect on the phenomenon. Researchers can attribute the change in the phenomena that occurred when the independent variable was manipulated, either across different groups or across trials of the same group, to the independent variable. Effectively, researchers can conclude that the independent variable is the mechanism operating behind the phenomenon since everything else in the research context, once specified was held in constant, across groups or trials, when the effect was found.

Although powerful, experiments are not bulletproof. One specific concern is the possibility that the phenomenon elicited in an experiment is determined primarily by the laboratory task itself, and thus is specific to that task. This is deeply problematic for any complex phenomena because the whole goal of these experiments is to understand and control phenomena in the world.

There are at least three cascading effects associated with the study of a lab-based phenomenon that does not generalize to the real world. First, the causes found to be linked to phenomenon in the context of an experiment may not play the same causal role for the phenomenon outside the experiment because the phenomenon elicited in the experiment, in the first place, is not representative of the phenomenon in the real world. If there is a disconnect between the behavior elicited in the lab and the behavior as it occurs in more real-world contexts, then the cause-effect relations shown in the laboratory would not operate in the real world. Second, if there is a disconnect, the wrong causal links established in the experiment between causes and effects will result in the proposal of wrong theories about how the world works and thus inspire studies that miss the point of research –learning about the phenomenon of interest in the real-world– but pursue research questions that will not eventually teach us about the world that we want to understand. Third, if there is a disconnect, researchers and the general public will build on those theories to create interventions that are not going to be helpful for the phenomenon of interest in the real-world, where interventions ought to work. Together these cascading effects demonstrate the far-reaching consequences that science's translation problem has for not only our generation of scientists and practitioners, but also for future generations.

Failures to translate findings obtained in experimental contexts to real-world contexts are a pervasive problem in science. In the lab, and in some people, the Bacillus Calmette-Guerin vaccine (BCG) for tuberculosis disease is highly effective. But the vaccine is also highly controversial and a clear example of complex multi-causal pathways that are not understood. The controversy over the BCG vaccine concerns the validity of experiments because they failed to show that the latitude of the region being tested interacted with the efficacy of the vaccine. Meta-analyses now estimate the protective efficacy varies from 0% to 80% in different populations and geographic regions, and it is still not clear whether the vaccine should be recommended in some parts of the world (Teo & Shingadia, 2006). The world –and how it affects this vaccine and tuberculosis—is complex and not understood. A similar example concerns the effectiveness of a surgery that supposedly reduced the incidence of cerebral infarction in patients diagnosed with the disease that is responsible for about a third of the new cases of strokes and is called asymptomatic carotid artery stenosis. The randomized trial that tested the effectiveness of the procedure was scientifically rigorous, however it was implemented in 39 clinical centers that were top-notch centers with highly trained doctors. The results of the study led to erroneously estimate the risk of the procedure that patients are exposed to when they are treated in other less specialized facilities that do not have highly trained doctors. Outside the experimental context defined in the original study, operative mortality increased 8-fold and the risk of stroke and death was about 3-fold higher (Rothwell, 2005). These translation problems occurred because science goes for the essential factors and imposes high control, and thus it does not attribute the results to the effect of other factors that are seemingly nonessential to the basic mechanisms tested but are life-threatening nonetheless. In another example, researchers have also failed to translate findings to the real world because the measure of interest that was used in the research context was not actually the measure that mattered outside the laboratory context. Fluoride was tested in the lab as a treatment to increase the bone density and in fact, the effect shown on bone density in the lab was robust. However, the fluoride treatment resulted in an increased risk of fractures when applied to patients (Riggs, Hodgson, O'fallon, Chao, Wahner, Muhs, Cedel & Melon, 1990). Bone density was a *surrogate* outcome, or a biological marker that was in theory an indirect measure of clinical outcomes; in this case, their chosen surrogate outcome was not linked to practical outcomes.

The Disconnect between experimental research and the real world is particularly problematic for the field of developmental psychology, the science that investigates how children develop. Development can be viewed as the product of changes occurring at nested timescales. Big developmental changes are seen at larger timescales of years and months, but smaller changes occurring day-in day-out one minute to the next one underlie these larger changes and thus it is the infant's every day behavior that brings developmental change (Smith & Thelen, 2003). In turn, the behavior that a child generates in real time is the product of the integration and interaction of the child's sub-systems and the characteristics of the context that comes in contact with the system (Smith & Thelen, 2003). As a result, it is paramount that experiments in developmental psychology understand the complex contexts –and underlying processes—that operate in the real world (or at least tell us something about the world), the context in which development takes place. There are a number of cases in the study of human development where highly replicated lab-based phenomena are later determined not to be what they initially seemed. One example is Mutual Exclusivity: infants map a novel name to a novel object in the context of a known object with a known name (Markman, Wasow & Hansen, 2003). This phenomenon generated a huge amount of interest in the 1980s as a "fast mapping" solution to word learning. But now we know that mutual exclusivity is unlikely to play a major role in the real world. Researchers have questioned that fast-mapping can explain the problem of word learning (Deák , 2000; Kucker, McMurray & Samuelson, 2015). Specifically, mutual exclusivity has been shown to depend on memory mechanisms and occur only when the potential referents are repeated over time (Mather & Plunkett, 2009). In the same way, the A-not-B task apparently measured object permanence, or the understanding that objects continue to exist even when they are not perceived, but in fact it is a lab-phenomenon that is completely task-dependent. Smith & Thelen

(2003) showed that young infants' errors of perseverating by reaching to the location without the object disappeared when small manipulations of the task were added. They proposed that the child's system self-organizes to produce cohesive patterns of action but that the system is also sensitive to initial conditions and interactions between the changing components of the child's system and the context. As a result, claims about development, its mechanism and causes, need to be based on experiments that elicit behavior in multiple and variable ways; behavior that can be generalized to the real world and does not depend on the arbitrary highly controlled tasks to be seen. The field should worry about how important a phenomenon or putative mechanism is if it cannot be measured in the complexity of everyday life.

The main goal of this paper is to understand a potential additional example of a *Disconnect problem: joint attention* when studied in discrete trial-based tasks and through observation of parent-infant free-flowing interactions.

The Phenomenon of Interest: joint attention

Interactions between two individuals or more often require the establishment of a common referent among the participants of the interaction. This referent could be the projector screen used by the speaker during a presentation. Individuals attend to the same thing at the same time by being sensitive to social cues that indicate an individual is engaged with an object including the individual's eye-gaze itself, gestures, speech, manual contact with an object or even the orientation of the body when directed to the attended object. When individuals pick up these cues and "follow" them by visually attending to the object of interest, they can build an interaction, or joint action, around the common referent. Specifically, by having their eye-balls on the same object they can identify the topic of the interaction and with additional cognitive processes that build on the establishment of a common referent such as language (and theory of mind as some would argue, Baron-Cohen & Cross, 1992), individuals can work together, teach or challenge each other, but also laugh, explain, create, solve, among others. Joint attention occurs when two individuals look to the same object at the same time and this everyday phenomenon is almost a prerequisite for the smooth dynamics of social interactions that occur every day in a variety of contexts.

Where does joint attention come from and how does it develop? Developmental psychologists trace the beginnings of joint attention to infancy. There are two dominant approaches to study joint attention between an infant and an adult; one elicits the behavior in the context of a trial-based experiment and one that elicits the behavior in the context of continuous free-flowing play sessions with objects. The two approaches conceptualize joint attention differently and they ask different questions (Tasker & Schmidt, 2008): the former defines it as (and examines) the *ability* of the infant to engage in joint attention given a standardized trial procedure that is constant across other infants. The latter defines it as (and examines) the emerging properties of the *moment* itself that is created by the joint effort of each infant and the parent throughout the course of a more naturalistic free play interaction. The two research contexts used to study joint attention not only conceptualize joint attention differently, a growing body of literature appears to suggest the phenomenon elicited by the trial-based procedure is not representative of the phenomenon that occurs in the more natural context of continuous free-flowing play sessions.

Joint Attention in the Context of Trial-based Experiments

In a typical trial-based experiment that measures Joint Attention (JA) skill, the infant is brought into the lab and sits on their parent's lap facing an unfamiliar experimenter who sits across the table in a clean room with only two objects on the table (or two posters on the wall). The experiment consists of between 3-8 trials and each trial lasts on average 5-8 seconds. Each trial begins with the experimenter turning the head and eyes to look at one of the two objects that are present in the room. The experimenter maintains the direction of gaze for a few seconds and then returns to look at the infant's face. Experimenter's speech is sometimes added to the main the "gaze cue" given to the infant in the form of the infant's name repeated three times (Mundy et al., 2003; Morales, Mundy & Rojas, 1998); experimenters however do not touch with their hands the object that is being cued with the experimenter's eye-gaze to the infant. Cameras around the room record the entire experiment so that trained coders can score the infant's ability to engage in Joint Attention. Even though joint attention in this context can technically occur for as long as the experimenter and infant look to the same object during the trial, the majority of scoring methods have measured the infant's ability to engage in JA based on whether the *first* look that the infant generated during the trial was directed to the experimenter-attended object (Brooks & Meltzoff, 2005; Morales, Mundy & Rojas, 1998; Mundy, Card & Fox, 2000; Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007; Markus, Mundy, Morales, Delgado & Yale, 2000). The most common first-look based measure in the literature is the proportion of correct trials, which is obtained by dividing the number of trials in which the infant looked first to the experimenter-attended object, by the total number of trials that the infant completed. Even if an infant spent the majority of the trial looking at the object attended-to by the experimenter, the infant's score in JA depends only on the first look.

The research studies that have used the trial-based task to elicit JA and have used the proportion of trials that were correct to measure JA ability, claim that infant's abilities to engage in Joint Attention are supported by the infant's abilities to do "gaze-following" because the only available cue to establish a common referent with the experimenter in the experimental context is the experimenter's eye-gaze and infants still engage in JA (Brooks & Meltzoff, 2005). Researchers who use to this trial-based approach to study JA have attributed the ability to engage in Joint Attention to an intention-reading module that takes information from the eyes and makes inferences about the intention of the other person. This literature also suggests that the ability to follow gaze is present by 6 months of age (Morales, Mundy & Rojas, 1998) and that at 9 months, infants generate a correct first look in 25% of the trials (Mundy et al., 2007). Different studies show that the gaze-following ability that supports JA undergoes dramatic change throughout the second year of life (Mundy et al., 2007; Vaughan, Mundy, Block, Burnette, Delgado, Gomez, Meyer, Neal & Pomares, 2003). Additionally, the evidence suggests the infant's ability to engage in joint attention through following (measured by the gaze-following task just described) or leading (a type of joint attention moment that was not even possible in the gaze-following task) is not the same and in fact recruits different brain networks (Mundy et al., 2007; Mundy, Card & Fox, 2000).

The claims made by the trial-based approach (with respect to how JA occurs, how it develops, how it relates to other infant abilities) hinge on their methods, as any claims do. After all, the claims come from results, results come from scoring the behavior, and behavior comes from the research context defined to elicit the behavior. The trial-based experimental approach creates a clean and controlled "social interaction" that is held constant across infants in which infants are tested for their ability to establish a common referent with an adult with the use of first-look based measures. In particular, the interaction that occurs is set up to take unique values with respect to who is the adult that is interacting and how she/he behaves, the type of cue given and the number of times is given, the amount of time between the cues given, the duration of the cue as well as the location of the objects when the infant was asked to jointly attend to the experimenter-attended object. These task characteristics once defined they are maintained equal across participants and thus provide a fair estimate of each infant's ability *under* the research context. But is the "social interaction" recreated by the research context and used to elicit an infant's ability to engage in JA representative of the interactions that infants partake of in their everyday lives? How reliable and valid are the estimates obtained by first-look based measures?

Coordinated Joint Attention in the Context of Continuous Free Play Sessions

Parent-infant free play sessions carried out in the laboratory are one excellent approximation to the play interactions of infants and their parents that take place in real-world contexts such as the home (Tamis-LeMonda et al., 2017). As a result, the context of a continuous free play session has been used by several research laboratories to elicit *Coordinated Joint Attention* between infants and their parents. When compared to the context of discrete trial-based experiments, this more naturalistic context allows us to determine if the behavior elicited in discrete trial-based is representative of infant behavior in everyday contexts. The term ***Coordinated Joint Attention*** **(Coordinated JA)** is used in this paper to refer to joint attention occurring in the context of Continuous Free Play sessions and to be distinguished from the phenomenon elicited in the Discrete trial-based experiment referred to as ***Joint Attention*** **(JA)**. The term ***joint attention*** without any capital letters is used to refer to the concept of the phenomenon without regards to the methodology and definition used by researchers to study it.

In a typical free play session that measures Coordinated JA moments, the parent and infant are brought into the lab and given a set of toys to play with for 5 or 10 minutes. Infants are sometimes sitting at a table across the parent, or are sometimes on the floor with the parent. The parent is told to interact and play with the infant, as they would normally do at home and thus parents are free to touch objects with their hands and to also talk while they play. Cameras record the interaction to measure both infant and parent behavior moment-to-moment capturing what the infant looked at, what the parent looked at, touched and said. Several recent studies have implemented head-mounted eye-trackers (Yu & Smith, 2013; 2017a; 2017b), which are attached to hats worn by the parent and the infant. By measuring where the parent and infant look at during the interaction, researchers define episodes of Coordinated Joint Attention. Some researchers define Coordinated JA as the moments in which the infant and the parent look to the same object at the same time (Yu & Smith, 2013) during the entire interaction and for as long as it happens, while some others define it as the moments in which the infant and the parent look to the same object while also looking to each other's faces (Tomasello & Farrar, 1986; Tomasello & Todd, 1983) during the entire interaction.

As it is clear, the context of free play session is designed to be as natural as possible in order to examine the emergence of Coordinated Joint Attention from the dynamic interaction that unfolds between the parent and the infant. In fact the only fixed characteristics of the "social interaction" are the objects they play with, the duration of the play session and whether they sit at a table or on the floor. Importantly free play sessions are representative of the interactions that infants are exposed to day-in day-out (Tamis-LeMonda et al., 2017) and that ultimately build the fundamental skills of the developing infant (Smith & Thelen, 2003) that support among others, the ability to jointly attend to objects with another individual. As a result, the continuous free play sessions provide researchers with the opportunity to assess the training ground that supports infant's development of Coordinated Joint Attention (as well as the development of infant's reaching, hand-eye coordination, attentional control, language). We turn to describe the Disconnect hypothesis for the study of joint attention by comparing the "social interaction" in which joint attention occurs during discrete-trial context and during the continuous free play context.

JA in the Context of Trial-based Experiments *differs* from Coordinated JA in Continuous Free Play Sessions

The existing evidence suggests that the phenomenon of Joint Attention recreated in the discrete trial-based experiment differs considerably from the phenomenon recreated in the continuous free-flowing play in at least five ways.

First, in the trial-based experiment, Joint Attention occurs in discrete units and is not immersed in the continuous and free flowing dynamics of human social interactions. In the real world, the coordinated action of looking together to an object emerges from an ongoing interaction that unfolds in time and is not defined arbitrarily by the onset of a trial, which a computer can present and trained experimenters too when instructed to do so. Research suggests that the discreteness of "fixed trial procedures" affects visual attention (Oakes, 2015), information processing (Colombo & Mitchell, 1990) and perhaps memory as well because memory is affected by the ways in which information is presented (Mather & Plunkett, 2011). Moreover, discrete trial-experiments define the timing of the trials arbitrarily and not based on how the adult adapts to the infant's state. In the discrete trial-based experiment, infants are asked to act regardless of their individual state. While this may be ideal to test all infants fairly (achieving good internal validity), it may not provide an accurate and ecologically valid assessment of an infant's ability to engage in joint attention with a social partner that *adjusts* in real time to the infant's state (Yu & Smith, 2013; 2017a; 2017b), which is the information that can be obtained during continuous free play sessions and that actually matters when we study how an infant engages in joint attention in everyday contexts.

Second, Joint Attention as it occurs in the social interaction recreated in the discrete trial-based task occurs only when the infant visually attends to the object attended to by the experimenter. In the real world, Coordinated Joint Attention results from the joint effort of parent and infant and thus infants play a dual role since they only "follow" but also "lead" the episodes of Coordinated JA by looking first to an object and providing the cues for the parent to follow "infant-led" episodes of Coordinated JA. Studies with 12-month-olds estimate infants are equally likely to lead than to follow since there is no reliable difference between number of parent-led and infant-led Coordinated JA moments (Yu & Smith, 2013; 2017a). The natural rhythm of equal following and leading could be a fundamental component that defines an interaction with an adult and thus influences infant's ability to follow the parent's lead. Without the opportunity to lead a joint attention moment, the infant may be losing contingencies that are part of a natural interaction between infants and adults and have been shown to influence infant learning and development (Gros-Louis, West, Goldstein & King, 2006; Miller, Ables, King & West, 2009).

Third, the trial-based experiment strips away the social cues that typically accompany an adult's engagement with an object in the real world. The discrete-trial experiment only provides a gaze cue for infants to determine the adult's engagement and thus infants can only be successful in this task by attending to the experimenter's face and following gaze. In continuous free play sessions, parents use other redundant cues to engage with objects including talk about the object and manual contact with the object (Chapter 3) and this is consistent with previous work (Bakeman & Adamson, 1984; Tomasello & Farrar, 1986; Yu & Smith, 2012; Yu & Smith, 2013). The evidence suggests that the only cue provided during discrete trial-based experiments, eye-gaze, is not used by infants, who are between 9- and 18-months of age, to engage in Coordinated JA. There is strong evidence to suggest that infants rarely look to the parent's face during play to perceive the parent's eye-gaze (Bakeman & Adamson, 1984; Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Suarez-Rivera, Smith & Yu, 2019; Yoshida & Smith, 2008; Yu & Smith, 2013) but follow the hands of the parents to engage in Coordinated JA (Yu & Smith, 2013; 2017a). As a result, gaze-following is not the most common pathway used by infants at this age to engage in Coordinated JA (Yu & Smith, 2017a). Gaze-following in fact, has been shown to fail in contexts that are more ambiguous even for adults (Corkum & Moore, 1998; Doherty, Anderson, & Howieson, 2009; Farroni, Johnson, Brockbank, & Simion, 2000; Langton, Watt, & Bruce, 2000; Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Vida & Maurer, 2012a, 2012b, 2012c). Outside the clean and limited context that is created by the discrete trial-based experiments to measure JA, gaze-following does not allow individuals to look to the same thing looked to by another individual.

Fourth, trial-based experiments for JA break the infant's own natural dynamics of their bodies since they are not able to engage with the objects while using their whole body. In meaningful contexts representative of an infant's life, hands and eyes are coordinated moving from one object to the next one (Yu & Smith, 2017b). Infants attend to the object visually but they also use their hands to explore the attended-object (Ruff, 1986). When infants are observed in real-world interactions, their attention is sustained and extended over time when children hold the object manually and rotate it (Ruff & Lawson, 1990). In particular, infants with tight child eye-hand coordination are those who are engage in more Coordinated Joint Attention (Yu & Smith, 2017b). When infants are allowed to touch the objects in company of a social partner and in the context of play, they devote the majority of time visually inspecting the objects and not the faces of those who are around and are not trying to get the child to look at a distant object but to support the child's attention and play. The discrete trial-based experiments may not elicit infant behavior that is representative of every day social interactions because they break the multimodal nature of object exploration that results in coordinated action with social partners.

Finally, one key difference between the phenomenon of joint attention defined under a discrete trial-based context and a context of continuous free play is the emphasis put on the orienting of attention to a social partner's cue but not the maintenance of that attention. There is only one solution to the task, and thus the infant can find that solution or not find it, but the infant cannot show the scientist that they are much better at coordinating attention in some other way, in an alternative solution not available in the discrete trial-based experiment. In the trial-based experiment, Joint Attention depends on measures that score the first infant's look to the experimenter-attended object. More real-world contexts do not necessarily impose those biases to the first responses and single gaze-following solution because infants take information from the world in all of their looks and may figure out the object that the adult attends to by generating additional looks to other objects in the environment including the hands of the adult. In fact, the relevant behavior that is needed in the real world is the continued engagement with the jointly-attended object to build the interaction, learn, discuss, create, etc. In free flowing contexts for the phenomenon, Coordinated Joint Attention is measured as the total amount of time looking at the same object (Yu & Smith, 2013), conceptually described here as an *all-trial-based* measures. Are first-look-based measures representative of the total amount of time in which joint engagement occurs between that infant and the experimenter during the entire trial of the experiment? The phenomenon captured during discrete trial-based experiments focuses on only the first looks generated and the degree to which this initial response is representative of the maintenance is an open question.

The five key differences between the discrete trial-based context and the continuous free play sessions raise the serious possibility that the phenomenon elicited in the laboratory experiment is not representative of the phenomenon in the real world -- contexts in which joint attention emerges naturally between infants and other adults around them.

Two previous studies hint there is disconnect between discrete trial measures of gaze following and coordinated social attention between parents and infants. Vaughan et al., (2003) measured 9-month-olds' joint attention skills using continuous parent-infant play sessions and a discrete trial-based experiment. In the former, infants and their parents played for 7 minutes with a set of toys provided by the experimenters and Coordinated Joint Attention was coded according to the coding scheme of Bakeman & Adamson (1984) and Tomasello & Farrar (1986), and thus represented moments of parent-infant engagement with an object and with each other's faces. Joint Attention in the context of a standard discrete trial-based experiment was measured as the percentage of trials in which the infant followed the gaze cue correctly in their first look. The results showed that Joint Attention was not correlated with any measure of Coordinated Joint Attention that included looks to the parent's face during play. Even though researchers used a coding scheme that counted Coordinated Joint Attention if there was gaze-following, researchers failed to identify the Disconnect problem in the phenomenon itself and argued instead that discrete trial-based experiments carried out by an experimenter are practically a better measure of children's abilities because "the variance in the social partner's (the experimenter) behavior is standardized, perhaps yielding a clearer picture of differences in joint attention skills among infants" (Vaughan et al., 2003, pp. 606). Researchers proposed that in continuous free play sessions between the infant and the parent it is difficult to tease apart the variance in infant from the variance in parent behavior, which given the multi-causal nature of free-flowing social interactions is likely correct. But if the discrete trial task does not relate to real world social interactions, what is it telling us?

Morales, Mundy, Crowson, Neal & Delgado (2005) reported findings that suggest the phenomenon elicited in discrete trial-based experiments is not representative of the *future* social abilities of infants in continuous free play sessions. At 6 months, infants' Joint Attention abilities were measured using a standard discrete trial-based experiment. At 24 months, Coordinated Joint Attention was measured in the context of continuous free play as the number of episodes in which the child and parent looked to the same object for at least 2 seconds while looking to the parent's face and talking about the object. The results showed the individual differences in Joint Attention at 6 months were not related to the individual differences in Coordinated Joint Attention that included infant looks to the parent's face at 24 months. These previous studies are consistent with the idea that measures of joint attention obtained in discrete trial-based experiments are not related to measures obtained in more everyday play contexts.

One possibility is that the two prior studies found a disconnect between contexts of joint attention because they scored Coordinated Joint Attention with coding schemes that do not capture typical infant behavior when looking to an object with the parent during play. According to the coding schemes used in these previous studies, a joint attention moment was only counted if the infant alternated gaze between the jointly-attended object and the parent's face. Nonetheless, the evidence strongly suggests that infants only look to the parent's face during play rarely (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013) and in fact the pathway of looking at the parent's face and then at the object is not common for infants this age (Yu & Smith, 2017b). Research that scores Coordinated JA even when the infant did not look at the parent's face but only focused on the jointly-attended object is needed to examine whether individual differences in JA and Coordinated JA with and without infant looks to the parent's face relate to each other.

A Rationale for the Current Study

A study that addresses the hypothesized joint attention Disconnect between discrete-trial experiments and continuous free play session is needed because of four main reasons: the context of free play sessions provides rich and valid information about how infants engage in Coordinated Joint Attention in the world, both approaches were designed to study the same phenomenon, both approaches predict language development and the discrete-trial based experiment is an excellent diagnostic tool.

Research is needed to examine the phenomenon of Coordinated JA as it occurs in the context of a continuous free play session because this is the context that describes how infants engage in joint attention with mature social partners in the real world. A close examination of these moments with state of the art head-mounted eye-tracking, can detail the properties that make up real-world joint attention and highlight the role of hands for the joint engagement with an object. Moreover, an investigation of the temporal properties of the moments of Coordinated JA can describe the amount of time that it takes for the parent and infant to join each other and whether the parent is playing an equal role for the leading the attainment of joint attention during 9-month-olds' play. Finally, this context also provides measures of other more general infant sensory-motor skills that are important for real-world learning and development.

Is Coordinated JA in continuous free play sessions, the behavior that is representative of infants' everyday lives, related to the JA abilities demonstrated in a discrete-trial based experiment? Researchers continue to use the contexts and measures of JA and Coordinated JA interchangeably. In fact, the consensus assumption is that the phenomenon being elicited across contexts is the same because it taps the same underlying infant abilities and it was designed to capture the same phenomenon. Nonetheless, there are clear differences between the approaches and in fact the measures used in the literature are strikingly different. What does it mean to engage in joint attention? It is unclear whether measures of JA even within discrete trial-based experiment capture variance when they measure the infants' first response to the cue as opposed the infants' all-trial response, and therefore studies that examine relation between all-trial- and first-look based measures are needed. Moreover, is the holding cue provided by parents during continuous free play session a critical component that could lead to JA if it were to be added by the experimenter to the gaze cue in the context of discrete trial-based experiments? It is possible that the addition of a holding cue could lead to more JA and that there is a Disconnect between the phenomenon studied in the lab and the phenomenon that occurs in the real world, but research is needed to address these possibilities directly.

Furthermore, measures of joint attention across contexts predict individual differences in language development. Infants between the ages of 9 to 12 months, who perform better in discrete trial-based tasks that measure first-look performance, are infants who have better vocabulary at 18 months (Brooks & Meltzoff, 2005) and at 24 months (Morales, Mundy & Rojas, 1998; Mundy et al., 2007; Markus, Mundy, Morales, Delgado & Yale, 2000). Similarly, individual differences in the amount of all-trial Coordinated JA obtained by a dyad during continuous free play sessions also predict or contribute to the prediction vocabulary at 15 months (Yu, Suanda & Smith, 2019) at 18 months (Smith, Adamson & Bakeman, 1998) and at 21 months (Tomasello & Farrar, 1986). Since both contexts for JA predict language development, one open question is whether the individual differences measured across contexts are related. Two measures may both predict the same outcome but be uncorrelated if each of the measures accounts for separate explainable variance in language development. As opposed to previous work that just related measures of JA to measures of Coordinated JA using less precise coding methods that emphasize infant looks to the parent's face, here we also tested relations between JA and other infant sensory-motor behaviors that support learning during play using state of the art techniques to measure Coordinated Joint Attention (both with and without infant looks to the parent's face), parent and infant behavior.

Finally, the discrete trial-based experiment is part of the instrument that clinicians use to diagnose Autism Spectrum Disorder (ASD). Children who fail to do gaze-following in the context of discrete trials with an adult in fact show one of the criteria used to diagnose ASD (Lord, Risi, Lambrecht, Cook, Leventhal, DiLavore, Pickles & Rutter, 2000). The discrete trial-based experiment is actually a very useful task for our society, as it captures meaningful individual differences and research is needed to find out the infant and parent behaviors that are predicted by the task. Here we relate the diagnostic measures of discrete trial-based experiments with the measures of Coordinated JA and measures of more general sensory-motor behaviors that are practiced by infants in their everyday. This research can tell if the diagnostic tool relates in any way to behavior representative of the real world.

For all these reasons, the overarching goal of this work is to assess how predictive are measures of JA in discrete trial-based experiments of infant behavior that occurs in the context of continuous free play, the context that supports infant development (Smith & Thelen, 2003). Even though continuous free play may help us learn more about real world joint attention and the properties of the phenomenon, both tasks were designed to measure joint attention and they both are predictive of the same outcome. Here we show strong evidence that infant behavior in the discrete trial-based context is not related to infant behaviors examined during continuous free play sessions. We propose that researchers need to move away from the discrete trial-based experiments if they want to understand the phenomenon of joint attention as it occurs in the real world.

The Current Study

In the present study, 9-month-olds participated in both a discrete trial-based procedure to measure JA and in a continuous free play session with the parent to measure Coordinated JA. Both tasks were designed to be exactly like the tasks used in the literature, the former with the parent and the latter with the experimenter in order to test the Disconnect between the approaches to joint attention that are used in the literature, in the exact same way they are being used.

The discrete trial-based experiment was carried out by an experimenter who shifted gaze to an object in 12 trials. Six Gaze trials had only a gaze cue while six other trials had a gaze in addition to a hold cue. The additional Gaze-Hold trials were created to measure the infant's ability in a context that is more similar to the natural interactions between infants and adults because the experimenter not only looked to the cued object but also touched the object. The continuous free play session lasted 4 minutes and took place at a table top between that infant and the parent. Dual head-mounted eye-trackers were used to code with objectiveness and precision both infant and parent's eye-gaze, and cameras around the room recorded parents' manual contact with the objects and the parents' speech during play.

This dissertation reports several results examining the relation of infant behavior elicited by the discrete-trial experiment and by the continuous free play session. There is strong evidence to suggest there is no relation across contexts and this dissertation reports it all. Several measures were obtained from the discrete trial-based experiment (both first-look and all-trial based measures) in the traditional and modified trial type to predict various aspects of infant behavior obtained from the continuous free play context. Consistent with Disconnect hypothesis, for all measures chosen from Gaze and Gaze-Hold trials and for all the tests conducted, the results show that the behavior elicited during discrete trial-based experiments is not related to the behavior elicited in the context of continuous free play.

Method

Participants

The participants were 27 infants (15 females) aged 9 months (MEAN=9.34, SD=0.21) and their parents who completed two visits. Visit 1 consisted of a discrete trial-based procedure and Visit 2 of continuous free play session. Visit 2 occurred only a week apart from Visit 1 and all 27 infants contributed data during both visits. Eight other infants were recruited but did not contribute data because they refused to wear the head-mounted device in Visit 2 or other technical failure occurred. All participating infants were given a small toy as compensation. The entire sample of infants consisted of predominantly working- and middle-class families. This research project was approved by the Institutional Review Board at Indiana University (protocol number 0808000094) and titled "Multimodal word learning".

Stimuli

A pool of 24 novel objects (on average, about 5 cm ^{x} 4 cm ^{x} 3 cm) was created for the discrete trial-based procedure at Visit 1. Each object appeared only once and objects were paired in sets of two. Six common toys, such as car and duck, (on average, about 9.50 cm ^{x} 6.5 cm ^{x} 5.0 cm) were used for the Continuous Free Play Session at Visit 2. Infants played with the six toys in 2 sets of three, and in each set one object was painted blue, one red and one green. All objects used in Visit 1 and Visit 2 were designed to be visually and manually engaging.

Experimental Room

At Visit 1, the infant sat on the parent's lap at a table (dimensions 151 cm ^{x} 76 cm ^{x} 66 cm) facing an experimenter. The experimenter sitting in front was approximately 75 cm from the center of the table and the child was 63 cm. During each of the trials, a wooden tray was placed in front of the infant on the table by the experimenter with the two objects used for each trial on each of the extremes of the tray. A camera was directed at the center of the infant's face and gaze direction was coded frame-by-frame from this recording, a reliable procedure widely used in many studies (e.g., Miller et al., 2009). Two additional video recorders also recorded the discrete trial-based procedure that was implemented to measure infants' JA abilities from a bird's eye view and a side view.

At Visit 2, the parent and infant sat at a small table (dimensions 61 cm ^{x} 91 cm ^{x} 64 cm) facing each other; the center of the tabletop was approximately 46 cm away from the infant's eyes and the parent's eyes. Both participants wore head-mounted eye trackers (Positive Science LLC, http://www.positivescience.com; also see Franchak & Adolph, 2010; Franchak et al., 2011) during the play session. The eye-tracking system attached to the infant's head through Velcro sewed into a hat and thus remained stable throughout the free play session. The eye-tracking systems for both the infant and parent included an infrared camera that recorded eye images – mounted on the head and pointed to the right eye of the participant– and a scene camera that captured the events from the participant's egocentric perspective. The scene camera's visual field was 108 degrees, capturing a broad view although less than the full visual field. Each eye tracking system recorded both the egocentric-view video and eye-in-head position (x and y) in the captured scene at a sampling rate of 30 Hz. In order to allow for automatic detection of the objects on the table, everything in the room –other than the objects and the hands and faces of the participants- was white. This environment elicited natural play centered on the toys on the table by the participating dyads. Three additional cameras recorded the interaction from third-person views.

Procedure

The procedure at Visit 1 consisted of a discrete trial-based experiment that had 12 trials. The same experimenter tested all infants in the sample. In each trial, the experimenter paired two of the novel objects and placed them on the tray. Each trial began with the experimenter looking to the infant's face and then performing a "social cue" directed to only one of the two objects on the tray for on average 4.90 seconds (SD=0.94). When the cue ended, the researcher looked to the infant's face and prompted the infant to reach for the objects by pushing forward the tray containing the objects. Infants were allowed to touch the objects until the experimenter collected the objects and moved to the next trial. The trials were randomized and two orders were created. Infants were randomly assigned to follow one of the two orders.

There were two types of trials and they differed by the type of cue performed by the experimenter. All infants completed both types of trial in a within-subjects design and thus each infant was asked to complete six Gaze trials and six Gaze-Hold trials. Gaze trials, were exactly like those used by researchers to measure JA skill (e.g., Brooks & Meltzoff, 2005). In a Gaze trial, the experimenter shifted the head and eyes to attend to one of the two objects on the table. Half of the six Gaze trials were completely silent and in half the experimenter provided a novel word inserted (X) in the carrier “Look X, X”. Both kinds of Gaze trials were combined for the analyses reported here because both showed similar of infants’ eye-gaze to the objects and to the experimenter’s face (Supplemental Material). In a Gaze-Hold trial, the experimenter not only shifted her head and eyes to attend to one of the objects on the table but also used her hands to pick up the object while looking at it. Gaze-Hold trials were designed to measure infants’ JA skill under a research context that allowed the experimenter to touch the objects and thus resembled more the context of free play between infants and adults. Half of the Gaze-Hold trials were silent and half were not. The results reported here combined both kinds of Gaze-Hold trials given that the presence of the carrier “Look X, X” did not have an effect on the distribution of infant’s eye-gaze (Supplemental Material). Silent and non-silent cues were only different in the duration of the cue itself. On average, Gaze cues were 4.59 seconds-long (SD=0.41, min=3.95, max=5.53) while Gaze-Hold cues were on average 5.21 seconds-long (SD=0.63, min=3.63, max=6.32), and they were significantly different, $t(26)=-6.55$, $p<0.001$. Throughout the testing period, parents were instructed not to engage with the objects or infant.

The procedure at Visit 2 consisted of a continuous free play session. Upon entering the experimental room a second experimenter and the parent showed a highly interactive pop-up toy to the infant as the first experimenter placed the headgear on the infant. The infant’s eye-tracker placement occurred in one movement while the infant held the interactive toy with both hands, and the scene camera was adjusted to ensure that the toy being manipulated by the infant was in the center of the scene camera. Fifteen calibration points for the infant’s eye-tracker were obtained by directing the infant's eyes toward an attractive toy and a laser pointer on the table. After placing the parent's headgear, the experimenter asked the parent to look at the laser pointing on the table in various locations to obtain fifteen calibration points for the parent's eye-tracker. During this visit, parents were instructed to play with their infant with 3 toys at a time, as they would naturally do at home. Parent-infant dyads played in four 1-minute-long trials, using two different sets of 3 toys in an alternating fashion across the 4 trials. The duration of the trials was chosen so that infants remained engaged in play with limited off-task behavior throughout the experiment.

Coding

Discrete trial-based Experiment (Visit 1).

The infants' eye-gaze during the experiment and the trial's cue onset and offset were coded to then derive measures of JA in this discrete trial-based context for joint attention. In addition, the infants' manual contact with the objects when given the opportunity to hold the objects was coded to obtain measures of the infants' more general sensory-motor skills used during the task. All coding from the discrete trial-based experiment was carried out by trained coders using Datavyu (Team D, 2014) on a frame-by-frame basis.

Infants' eye-gaze. Infants' eye-gaze, captured by a camera directed at the children's face, was coded frame by frame by trained coders. The possible ROIs were the two objects on the tray, the experimenter's face, the center of the table and any other region around the room. Although infants' eye-gaze was coded for the entirety of the trial, infants only looked to the same object the experimenter looked to (and thus engaged in joint attention) during the social cue and thus this was the most relevant window to examine infants' eye-gaze. Three independent coders who coded infants' eye-gaze in three of the sessions assessed reliability of the coding. All codings were compared yielding on average 88.66% of agreement (SD=3.51, MIN=85, MAX=92).

Experimenter's cue. Trained coders annotated the times at which each cue performed by the experimenter started (onset) and ended (offset). The onset of the cue was defined as the moment in which the experimenter had her eyes on the target for Gaze trials, and as the moment in which the experimenter fixed both eyes and hands on the target for Gaze-Hold trials. The offset of the cue was defined when the experimenter stopped engaging with the object and returned to look to the infant's face. Three independent coders annotated three of the sessions to assess reliability. Coders agreed on average on 96.67% of their judgments (SD=3.06, MIN=94, MAX=100).

Infants' holding. Infants were allowed to touch the objects on the tray after each of the cues. Trained coders annotated frame by frame the objects that made contact with the infant's hands in addition to whether the hands were resting on the table or doing something else such as hands up in the air. Inter-rater reliability was calculated by the comparing the codings of three independent coders in 3 sessions. This comparison yielded agreement of the coders on average in 91.67% of their annotations (SD=7.64, MIN=85, MAX=100).

Continuous Free Play Session (Visit 2).

Measures of Coordinated Joint Attention in the continuous free play session (Visit 2) were obtained from the base coding of infant and parent eye-gaze. Other key parent and infant behaviors were coded from the play session including infant and parent manual holding of objects and parent's speech.

Infant and parent eye-gaze. For each frame, the eye-tracker system of the infant and the parent generated an image with a crosshair superimposed indicating infant and parent gaze, respectively, at that moment. Coders used these images to determine when gaze fell in each of four regions of interest (ROIs): the three objects in play at any time and the partner's face. Each ROI was strictly defined in terms of the in-view pixels belonging to the target. Inter-coder reliability was high and estimated to be the same as the reliability reported by previous published research using the same coding protocols (see Yu & Smith, 2012; 2013; 2017a; 2017b).

Infant and parent hand contact. Parent and infant manual contact with an object was coded from images captured by the overhead camera and the other two third-person cameras. Coders made objective frame-by-frame yes/no decisions that a hand was in contact with an object at that frame and did so for both hands for both the infant and the parent. Inter-coder reliability was high and estimated to be the same as the reliability reported by previous published research using the same coding protocols (see Yu & Smith, 2012; 2013; 2017a; 2017b).

Parent speech. The speech generated by parents during the continuous free play session was fully transcribed and divided into utterances. Each utterance was defined as a string of speech that contained vowel sounds and occurred between two periods of silence lasting at least 400 milliseconds (Pereira, Smith & Yu, 2014; Suanda, Smith & Yu, 2016; Yu & Smith, 2012). The criterion used to code the parent's talk excluded sounds that were coughs, raspberries or sighs produced by the parent. Parents' talk included labeling utterances –such as "it's a big tower"– but also utterances that are filler sounds such as "yeah" or "uh oh." Utterances that included at least one name of the objects played with were referred to as naming utterances.

Statistical Analyses

The results are organized in four different sections. Section I reports first-look and all-trial measures of JA in the context of trial-based experiments, relations between them and individual differences. This section tested whether first-look-based measures of JA are estimates representative of all-trial infants' JA. Section II reports measures of Coordinated JA in the context of the continuous free play session as well as individual differences. This section supported the Disconnect hypothesis because it shows that in real-world contexts infants also lead moments of joint attention and when they follow, they use additional cues that the parents used to engage in Coordinated JA and not just do gaze-following. Section III shows analyses that predict in total 16 infant behaviors that occurred during play from first-look and all-trial-based measures of JA performance in the discrete trial-based task. Section IV examines the temporal dynamics of the phenomenon as it is elicited in each context for joint attention.

Various statistical analyses were used in the study and they were all implemented in the R environment (version 3.3.2) (R Development Core Team, 2016). Simple paired t-tests were used to compare two means obtained from the subjects and were reported only when the data fit the assumptions of the test including normality of the differences between each subject's means. Correlation analyses were used to examine relations between two variables. Pearson coefficients were estimated only when both of the variables were normally distributed, had a linear relation and had equal variance, while non-parametric Spearman coefficients were estimated when the data did not meet these assumptions. Finally, multiple linear regression was used to predict a variable from more than one independent variables. The models reported here met all assumptions of multiple linear regression as any necessary corrective measures were carried out to fit model assumptions.

Results

Section I. JA in the Context of a trial-based Experiment

On average infants completed 6 Gaze trials (SD=0.48, min=5, max=7), which included only a gaze cue directed towards a target object, and 6 Gaze-Hold trials (SD=0.48, min=5, max=7), which included both a gaze and hold cue directed towards a target object. What does it mean to engage in joint attention in this context? Infants' ability to engage in Joint Attention during this task was measured with six different measures in order to follow both approaches to study joint attention during a "social interaction". *All-trial based Measures* 1-3 follow Yu & Smith (2013) and those who capture the infant's engagement with the object of adult's engagement *for as long as* the adult is engaged during a social interaction in the context of continuous free play. *First-look based Measures* 4-6 follow Mundy and colleagues (Brooks & Meltzoff, 2005; Mundy et al., 2007) scoring, which focuses on the infant's first look generated in response to the cue performed by the experimenter during a social interaction in the context of discrete trial-based experiments. Table 1 shows descriptive statistics of the all the measures for both Gaze-Hold and Gaze trials. According to all measures except the frequency per minute, infants engaged in more JA when the experimenter performed a Gaze-Hold cue. Moreover, the variability in the infants' abilities to engage in JA during Gaze trials was usually greater than the variability in the infant's abilities to engage in JA during Gaze-Hold trials. Before reporting the relevant individual differences that are picked up by the discrete trial-based experiment, we examine Measures 1-3 first, then Measures 4-6, and finally the relations between them in order to examine: first, the reliability of the first-look based measures when compared to all-trial based calculated for the same trial type; second, the reliability of the measures when compared to same measure across trial types.

Table 1.

Measures of Joint Attention in the Context of trial-based Experiment

	Gaze-Hold trials				**Gaze trials**			
	Mean	SD	Min	Max	Mean	SD	Min	Max
1. Prop. of time spent in JA	0.83	0.09	0.61	0.96	0.40	0.11	0.25	0.64
2. Frequency per min of JA episodes	16.22	3.79	11.10	24.02	18.55	5.02	10.37	29.02
3. Duration of JA episodes (s)	3.64	0.90	1.82	5.30	1.29	0.50	0.65	2.45
4. Prop. of trials that were correct (1st look-based)	0.67	0.17	0.33	1.00	0.50	0.24	0.17	1.00
5. Difference score	2.07	2.16	-2	7	-0.04	2.79	-4	6
6. Looking score	2.41	2.04	-1	7	0.67	2.59	-3	6

JA according to first-look and all-trial Measures 1-6.

We describe infants' performance in the discrete trial-based experiment according to both all-trial and first-look based measures starting with all-trial measures shown in Table 1. According to Measure 1, infants engaged in Joint Attention with the experimenter, on average in 0.83 (SE=0.02, SD=0.09) of the Gaze-Hold cue while they engaged in Joint Attention in 0.40 (SE=0.02, SD=0.11) of the Gaze cue. Figure 1 shows the average proportion of the cue time spent in JA, in gazing at the experimenter's face, at the distractor object or at the center of the table, calculated as an average of the subject averages. The figure shows that Joint Attention occurred more during the Gaze-Hold trials ($t(26)=20.34$, $p<0.001$; mean of differences=0.43, 95% confidence interval [0.39-0.48]) while infant looks to the experimenter's face occurred more during the Gaze trials (t(26)=-9.13, p<0.001). The overall increase in proportion of time spent in Joint Attention during Gaze-Hold was due to significantly longer infant looks to the target object (Measure 3) but relatively the same frequency of infant looks to the target (Measure 2). Infants' looks to the target object during the Gaze-Hold cue were on average 3.64 seconds-long (SE=0.17, SD=0.90), which is significantly longer than the infants' looks to the target object during the Gaze cue that were on average 1.29 seconds-long (SE=0.10, SD=0.50), t(26)=13.31, p<0.001. Joint Attention, created by infants looking at the target object, occurred on average 16.22 times per minute (SD=3.79) during the Gaze-Hold cue and 18.55 times per minute (SD=5.02) during the Gaze cue, t(26)=-1.77, p=0.09. The relation between all-trial Measures 1-3 was not plotted because mathematically, Measure 1 is computed by multiplying Measure 2 and 3, and thus they provide complementary information to characterize the overall proportion of time during the cue spent in JA. We used the measure of proportion of time during the cue (Measure 1) and not the other two all-trial measures for the remainder of the paper in order to capture infants' Joint Attention skills when JA during the whole trial is scored.

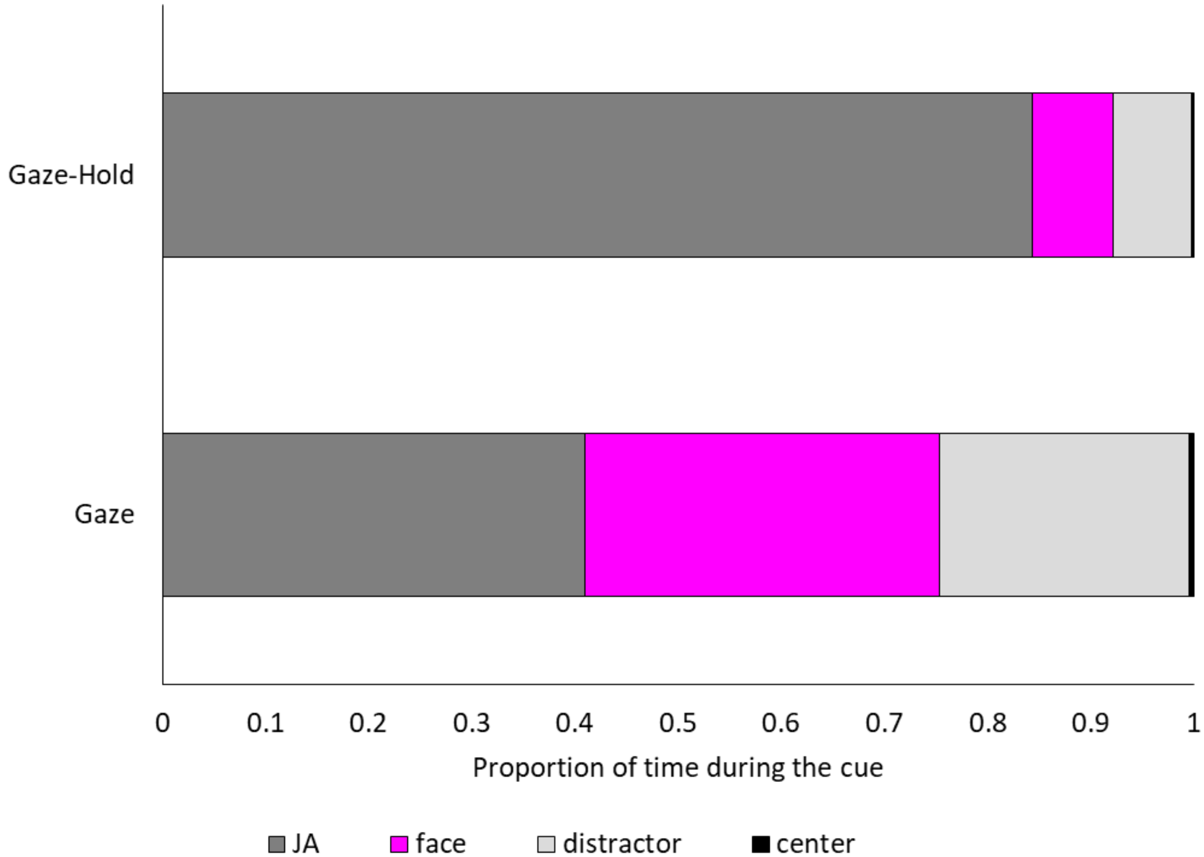


Figure 1. Mean proportion of time spent in Joint Attention and in infant looking at the face, distractor or center during the cue for Gaze-Hold and Gaze trials.

Infants' ability to engage in Joint Attention during the trial-based task was also described with Measures 4-6 shown in Table 1, which focuses on the infant's first response to the cue performed by the experimenter. Measure 4 was obtained by, after scoring each individual trial as correct or incorrect, taking the number of correct trials, for each individual infant, and dividing it by the total number of trials to obtain a proportion of trials that were correct. On average, infants had 0.67 Gaze-Hold correct trials (SE=0.03, SD=0.17) while they had 0.50 Gaze correct trials (SE=0.05, SD=0.24). There was still a significant difference in the proportion of correct trials between the two cue types ($t(26)=2.82$, $p=0.01$) but the effect size was smaller (mean of differences=0.17, 95% confidence interval [0.05-0.30]) than the size of the cue effect measured as the proportion of time in JA measure (Figure 1). Although some individual infants “failed” to follow the Gaze cues with their first look in most of the trials, the group as a whole followed the cue in half of the Gaze trials and this level of performance in the six Gaze trials was in fact higher than the performance of infants this age in a gaze-follow experiment (Mundy et al., 2007, about 25%).

In addition to the proportion of correct trials, we calculated first-look Measures 5-6 given the heterogeneity of first-look-based-measures utilized in the literature (Mundy et al., 2007; Brooks & Meltzoff, 2005; Morales, Mundy & Rojas, 1998; Mundy, Card & Fox, 2000; Markus, Mundy, Morales, Delgado & Yale, 2000). *The difference score* (Measure 5) subtracted the incorrect trials from the correct trials while *the looking score* added the scores given to all the trials for each infant (Measure 6). A score of 1 was given when the child’s first look was directed to that target cued object, while a score of -1 and 0 were given when the child’s first look was directed to the distractor object or somewhere else, respectively.

Figure 2 shows subject histogram of first-look Measures 4-6 obtained during Gaze-Hold along the diagonal. This figure also shows scatterplots between the measures as well as the Pearson correlation coefficients. The subject histograms show that difference scores were usually positive although a few were zero or negative (M=2.07, SD=2.16). Looking scores as well tended to be greater than zero, thus showing infants usually obtained a score of 1 in their trials (M=2.41, SD=2.04). These results show again that infants’ ability to engage in JA was high in Gaze-Hold trials even when each infant’s first look was used to measure performance. Moreover, Figure 2 shows that there are almost perfect correlations between the first-look based measures.

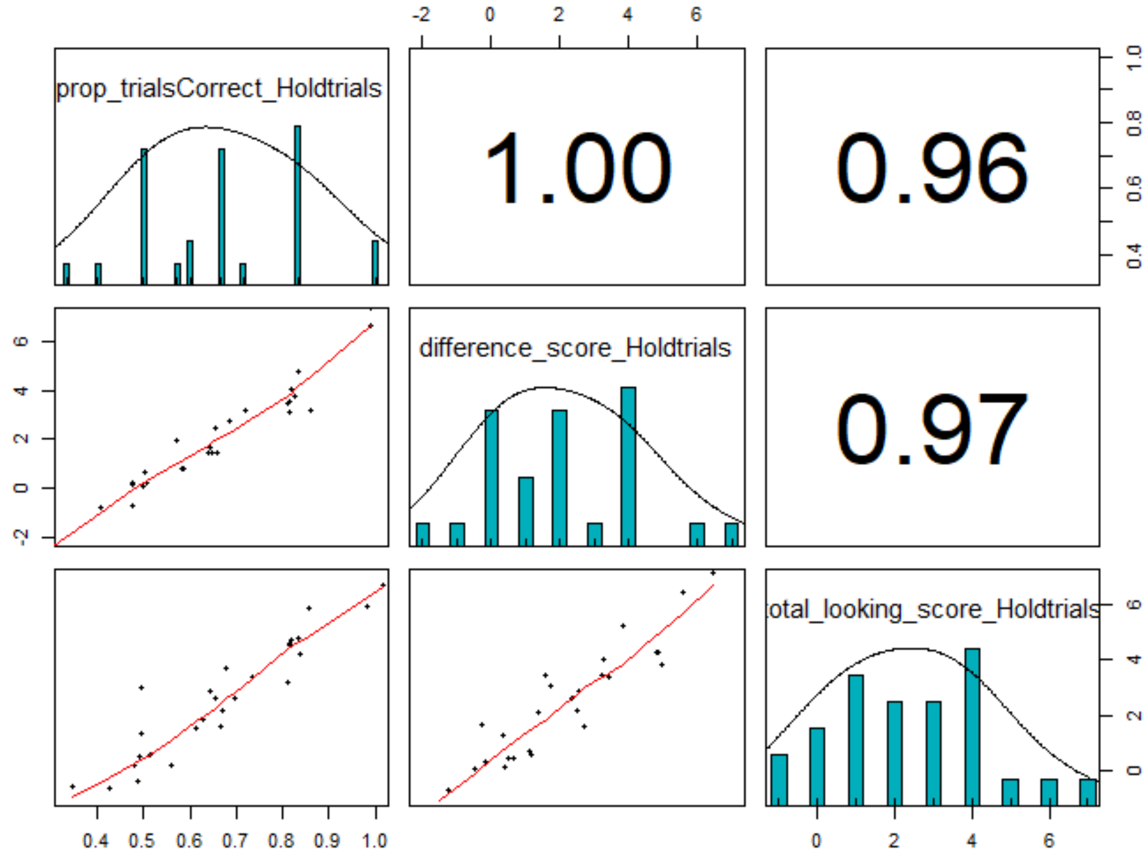


Figure 2. Histograms of Measures 4-6 and scatterplots (with a locally weighted smooth line LOESS- in red) between these literature-based measures for Gaze-Hold trials. Because all measures are normally distributed, the figure also shows the Pearson correlation coefficients between each pair of measures.

Figure 3 shows first-look Measures 4-6 during Gaze trials and the relations between them. As it is clear, the difference scores were skewed to the right with most infants showing low and negative difference scores while only a few had high difference scores (M=-0.04, SD=2.79). In the same way, looking scores were skewed to the right with most infants showing lower looking scores because their first look was directed to the experimenter's face or distractor object but not to the experimenter-attended object (M=0.67, SD=2.59). The distributions' skew present in Figure 3 but not in Figure 2, show that individual differences obtained from Gaze trials are more variable than individual differences obtained from Gaze-Hold trials. Finally the figure shows that there are almost perfect correlations between first-look Measures 4-6 for Gaze trials. Researchers that use Measures 4-6 to score infant JA in the context of Gaze trials are practically capturing the same variance and thus the heterogeneity in measuring JA in the discrete trial-based experiment may not be justified. Therefore, for the rest of the analyses reported in the paper only use the measure of proportion of correct trials (Measure 4) and not the other two first-look measures to capture infants' JA skills in the ways that previous studies have done.

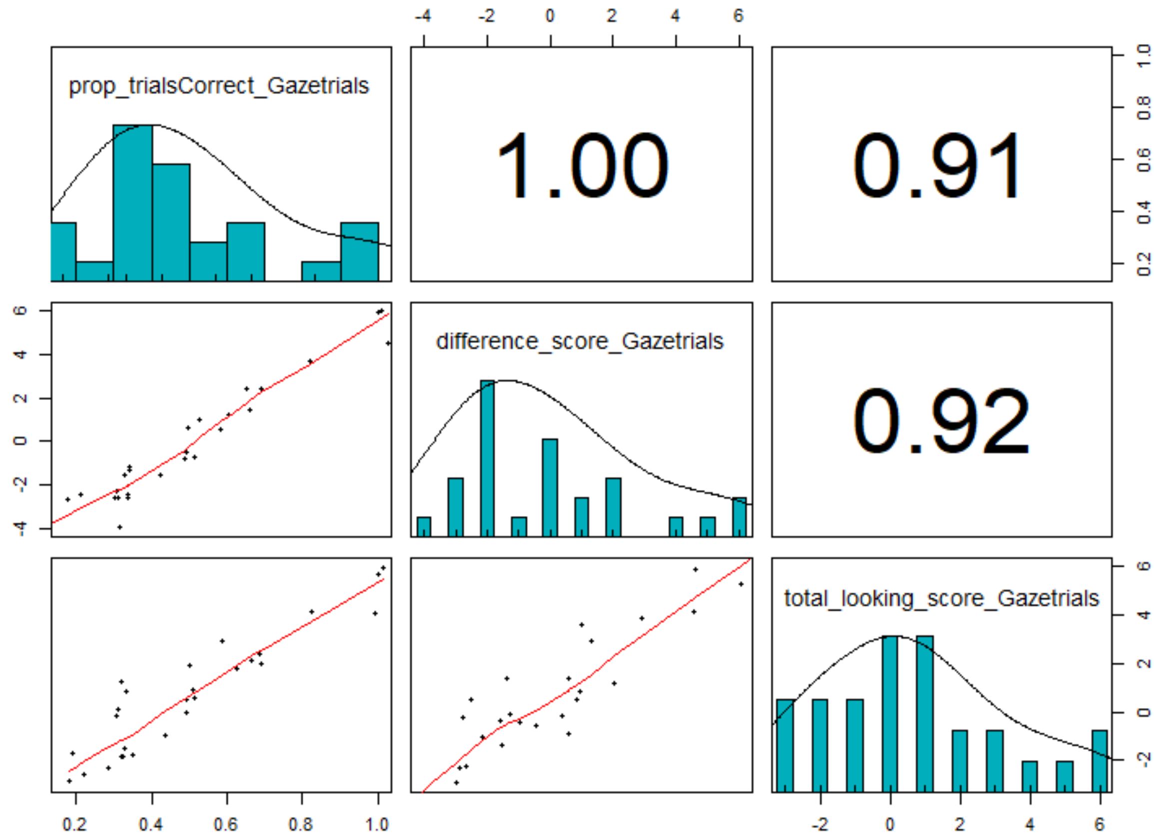


Figure 3. Histograms of measures 4-6 and scatterplots (with a locally weighted smooth line LOESS- in red) between the literature-based measures for Gaze trials. Because all measures are

not normally distributed, the figure also shows the non-parametric Spearman correlation coefficients between each pair of measures.

Together the results show that infant's abilities to engage in JA, according to Measures 1-6, were low and, according to Measures 4-6, more variable in Gaze trials, which are the trial type that is traditionally used to measure JA ability in the context of trial-based experiments. Moreover, even though researchers use different first-look-based measures to assess JA, all of the measures tap the same underlying sources of variance. We turn to examine whether the variance captured by the chosen first-look-based measure (Measure 4) is related to the other chosen all-trial measure (Measure 1) given the apparent disconnect between the measures themselves.

Correlations between all-trial and first-look measures.

What is the relation between the measures of Joint Attention that score the whole trial and measures that score the first look in discrete trials? Figure 4 shows correlation plots between Measure 1 and Measure 4 for Gaze-Hold trials and Gaze trials separately. The Pearson correlation is 0.27 (p=n.s) for the relation between measures of the Gaze-Hold trials. Due to non-normality, a Spearman correlation is -0.07 (p=n.s) for the relation between measures of the Gaze trials. This lack of correlation between Measure 1 and Measure 4 is problematic for the investigation of infant JA using discrete trial-based experiments because even when comparing measurements obtained from the *same* context for JA and trial type, there does not seem to be a reliable estimate of the infant's ability to engage in JA. The evidence suggests that the widely used measures in the literature that score infants' JA based on the first look are not representative of the infants' JA skills shown during the entire experiment.

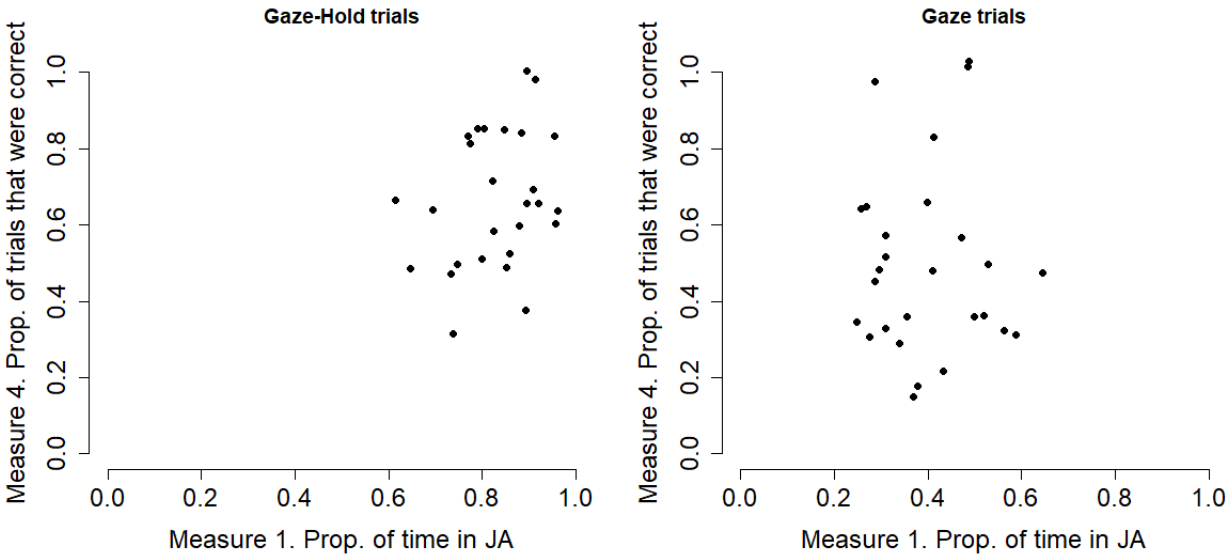


Figure 4. Mean proportion of trials that were correct (Measure 4) as a function of the mean proportion of time during the cue that was spent in JA (Measure 1) for Gaze-Hold (left) and Gaze trials (right).

Figure 4 shows that, especially for Gaze trials, there is not enough variability in the proportion of cue time spent in JA across trial types (Measure 1). Visual inspection of Figure 4 shows that the all-trial measure elicited less variability than the first-look measure, but we also compared the ratios between the infant with the highest and the lowest value according to Measure 1 and Measure 4 to show this. According to Measure 1, for Gaze trials, the infant with the highest proportion of time spent in JA engaged in 2.56 times more the proportion of time spent in JA by the infant with the lowest proportion of time spent in JA (0.64/0.25=2.56). This ratio is smaller than the ratio for Measure 4. According to Measure 4, for Gaze trials, the infant with highest proportion of trials that were correct had 5.88 times more the proportion of trials that were correct by the infant with the lowest proportion of trials that were correct (1.00/0.17=5.88). The ratio of Measure 4 doubles the ratio of Measure 1, and is one way to show that variability is reduced when infants' abilities are measured during the entire trial.

Correlation needs variability, and variability depends partially on sample size. First-look based measures may not represent all-trial behavior and increase variability in infants' responses because these first-look measures are based on a small sample of infants' behavior. As opposed to Measure 1 that sampled infants' JA behavior during the entire cue for *as long as* the experimenter attended to an object, Measure 4 only sampled few frames from each infant's eye-gaze behavior to produce a measure in which individual estimates are more likely to vary. Therefore, first-look based measures, such as Measures 4-6, may not be robust estimates of an infant's overall ability to engage in JA because they score the one shot establishment of JA when a cue starts.

Alternatively, it is possible that first-look Measure 4 are a reliable estimate of infants' JA abilities and in fact the ability to orient is just not related to the maintenance of JA, which may depend on other underlying processes and in fact just shows less variability because there is not much variability in the maintenance of JA. If so, then there should be correlations within Measure 4 across trial types because if Measure 4 is reliable and robust, then the orienting response in one type of trial should at least be related to the orienting response demonstrated in the different type of trial. We turn to examine the robustness of Measure 1 and Measure 4, separately, across trial types.

Correlations between Gaze-trials and Gaze-Hold trials.

Figure 5 (left) shows a scatterplot of Measure 1 obtained in Gaze-Hold trials as a function of Measure 1 obtained in Gaze trials. The measure of proportion of time spent in JA across trials had less variability but was still reliable across trial type. Individual differences in Joint Attention during the entire Gaze-Hold and Gaze trials were strongly correlated, $r_{Pearson}=0.43$, $p=0.02$. On the contrary, individual differences in Gaze-Hold trials were not correlated with the individual differences in Gaze trials according to the first response. Figure 5 (right) shows a scatterplot of Measure 4 obtained in Gaze-Hold trials as a function of Measure 4 obtained in Gaze trials. The results show that infants who engaged more in Joint Attention with their first look when given a Gaze-Hold were not the infants who engaged more in Joint Attention with their first look when given a Gaze cue, $r_{Spearman} = -0.13$, n.s.

These findings show that Measure 4 captured variance that is not related across trial types while Measure 1 did. The results could suggest the first look generated by the infant is not a valid and reliable measure of infants' JA all-trial ability because it is the same infant completing both trial types and they are not even related. Another possibility is that correlating Measure 4 obtained in Gaze trials with Measure 4 in Gaze-Hold trial is not the best test to examine robustness of the measure because the Gaze-Hold trial is a completely different context to elicit behavior. If that is the case, the results still support the main claim of the paper which is that the research context in the trial-based experiments elicits behavior that is very context-dependent and affects different infants in different ways even when one aspect of the context is changed. In this case, when hands were added to the social cue all infants performed better in following the cue but the ability shown by an infant during Gaze-Hold trials was not correlated to that infant's ability measured by trials without holding.

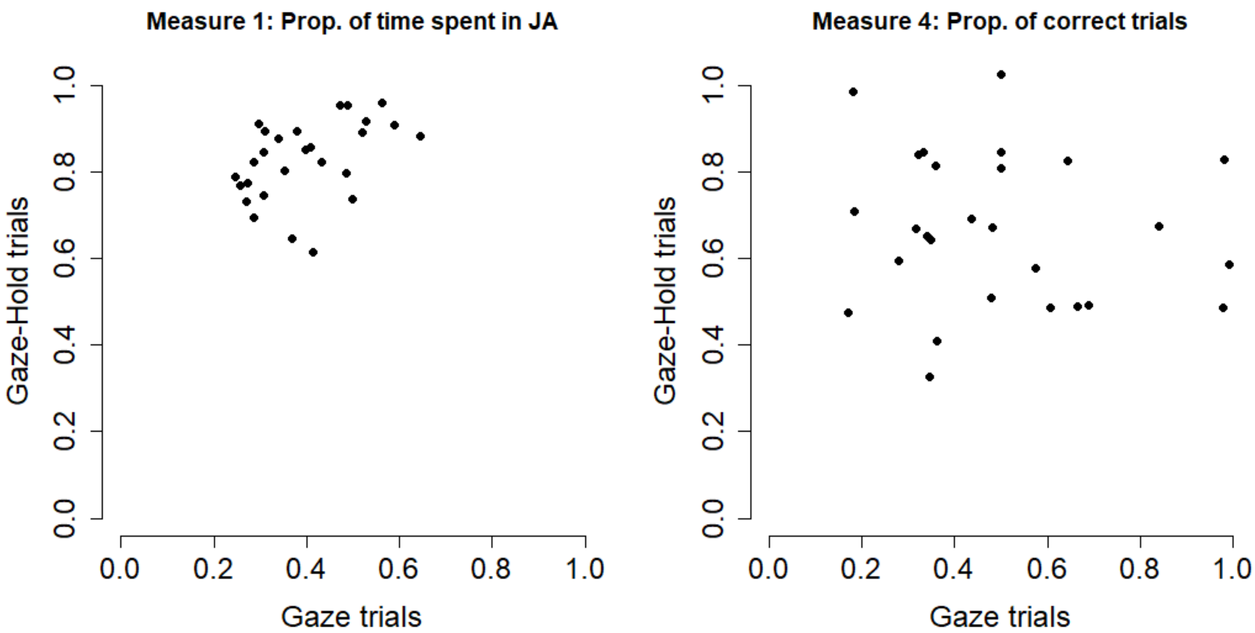


Figure 5. Performance in Gaze-Hold trials as a function of performance in Gaze trials for individual infants according to Measure 1 (left) and Measure 4 (right).

Individual differences in JA in the context of trial-based experiment.

Measures that use the first infant's look were not correlated with neither the infant's engagement in JA during the entire trial (Figure 4) nor the infant's first response to different trial types (Figure 5). Therefore, the individual differences in infants' performance in the discrete trial-based experiment should be described according to both first-look and all-trial measures. Figure 6 shows the individual differences for both Measure 1 and Measure 4, measures used in the next section to predict Coordinated JA.

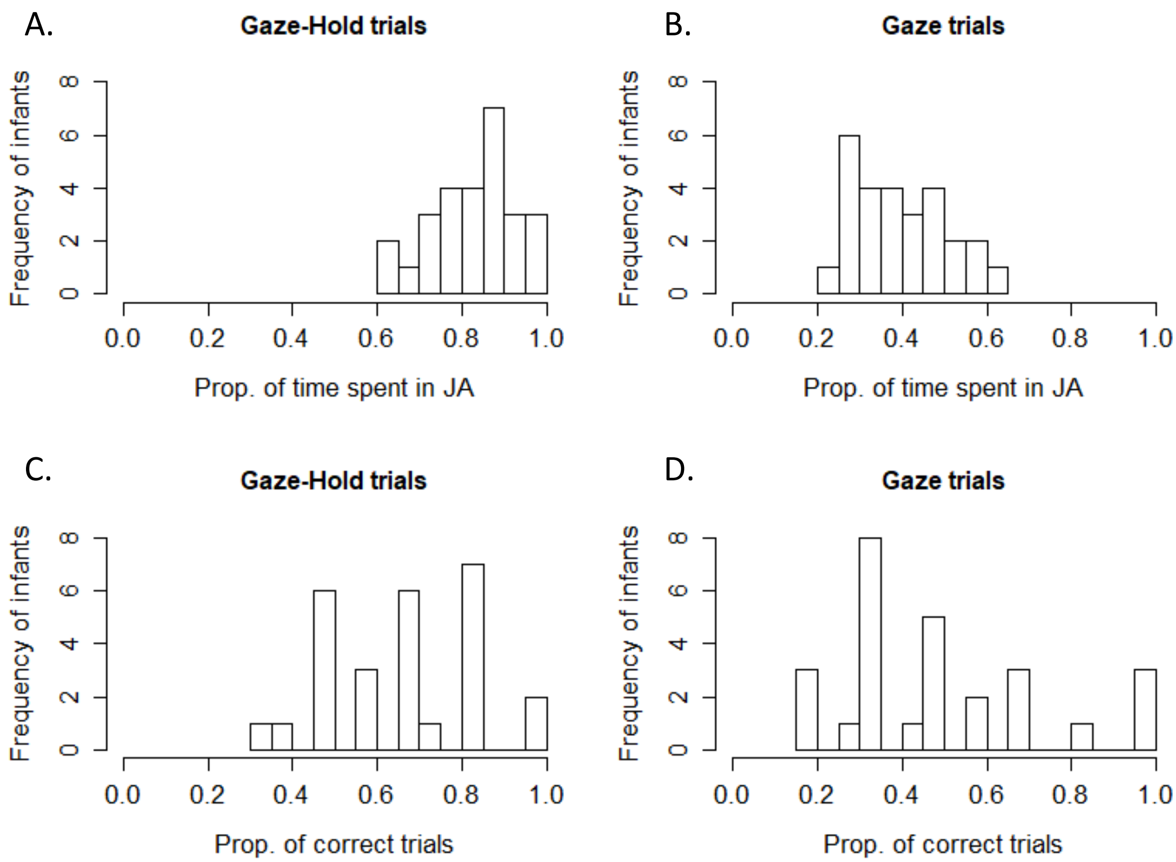


Figure 6. Individual differences in relevant measures of JA during the discrete trial-based experiment.

In this results section, infants' joint attention abilities were described in the context of discrete trials according to two different types of measures: measures based on the infant's engagement with the adult-attended object for as long as the adult attended and measures based on the infant's first response to the cue. Importantly, both types of measures used were not related to each other and thus predict infant behavior in the context of the free play session separately. However, there is evidence to suggest first-look-based measures, such as those used traditionally in JA experiments, are not reliable, robust estimates of all-trial infants JA abilities. The proportion of trials that were correct based on the first look was not related to the proportion of time during the cue in which the infant looked to the experimenter-attended object, a problematic finding for the investigation of joint attention using these measures in the context of discrete trial-based experiments. Moreover, the same first-look measure, proportion of Gaze trials that were correct, was also not related to the proportion of Gaze-Hold trials that were correct, but the proportion of cue time spent in JA was related across trials. Together, these results suggest that first-look based measures are highly variable and not valid estimates of neither all-trial JA nor, as we hypothesized, real-world infant behavior. Previous work that examines the test-retest reliability of gaze-following shows that first-look based measures are consistent for an infant when tested hours later in the same task with the same measure. However, the use of an all-trial based measure and the use of a different test target another important aspect of measurement, which is validity. Measurements need to be both reliable and valid to contribute to the investigation of joint attention.

Section II. Coordinated JA in the Context of Continuous Free Play

Parents and infants established common referent during the play interaction that occurred at Visit 2 by looking to the same object at the same time. Consistent with previous research using the free play session context to elicit the Coordinated Joint Attention, we first obtained measures of infant and parent's eye-gaze during play and then found the moments in which they both coincided on the same object at the same time. Dyads varied in the degree to which they coordinated visual attention during play, yielding meaningful individual differences. In this section, we report those individual differences with respect to Coordinated JA moments during play as well as to the different types of Coordinated JA moments.

Infants and parents looked to objects on the table, to each other faces and around the room during the free play session. On average, infants spent 0.56 (SD=0.15) of playtime looking at objects, while parents did so for 0.36 (SD=0.10) of playtime. Infants attended to the parent's face for 0.19 (SD=0.12) of the play session while parents looked much more to the infant's face. On average, parents looked to the infant's face for 0.49 (SD=0.14) of the play session. The remainder of the interaction was spent looking around the room (M=0.24, SD=0.11 for infants and M=0.15, SD=0.07 for parents). During play, infants generated on average 21.06 looks directed to objects per minute (SD=5.89) while parents generated on average 34.76 (SD=10.25). On average, infant looks to objects were 1.64-seconds-long (SD=0.48) while parent looks to objects were 0.64-seconds-long (SD=0.20). Infant and parent looks were used to code for moments of Coordinated Joint Attention during this natural free-flowing context of continuous parent-infant play. Coordinated Joint Attention was classified as parent-led or infant-led based on whether the parent or the infant was the first one to look to the to-be jointly-attended target. Furthermore, parent-led Coordinated Joint Attention was coded both when the infant looked to the parent's face (Gaze-following) as previous research did, but also when the infant did not look to the parent's face but instead followed the hands of the parent to engage in Coordinated JA (Hand-following). The following sections report individual differences for all of these levels of categorizing Coordinated JA.

Individual differences in Coordinated JA.

Coordinated Joint Attention was coded when the infant and parent's eye-gaze fell at the same time on the same object for at least 500 milliseconds. We report statistics on the proportion of playtime spent in Coordinated JA as well as on the frequency and duration of Coordinated JA moments. Table 2 includes the descriptives for all the measures of Coordinated JA that occurred during play.

Table 2.

Measures of Coordinated Joint Attention in the Context of Continuous Free Play

	Proportion of time				Frequency per minute				Mean duration			
	M	SD	Min	Max	M	SD	Min	Max	M	SD	Min	Max
Coordinated JA	0.21	0.09	0.04	0.39	6.52	2.26	2.31	9.57	1.88	0.49	0.94	2.95
Coordinated JA, parent-led	0.08	0.04	0.02	0.15	2.50	1.20	0.45	5.00	1.78	0.58	0.82	2.65
Coordinated JA, infant-led	0.12	0.06	0.02	0.26	3.84	1.51	1.15	7.12	1.71	0.62	0.58	3.39
Coordinated JA, parent-led through Gaze-follow	0.03	0.02	0	0.07	0.73	0.56	0	2.27	1.81	1.71	0	8.59
Coordinated JA, parent-led through Hand-follow	0.03	0.03	0	0.09	0.91	0.76	0	2.73	1.76	1.17	0	4.87

On average, dyads engaged in Coordinated Joint Attention for 0.21 (SD=0.09) of the play session. These bouts of Coordinated JA for a dyad were on average 1.88 seconds-long (SD=0.49) and they occurred on average 6.52 times per minute (SD=2.26). Mathematically, the proportion of time spent in Coordinated JA is computed by multiplying the frequency of Coordinated JA times the mean duration of Coordinated JA and dividing by 60 seconds, and thus the proportion of time spent in Coordinated JA represents both aspects and characterizes dyads' overall joint visual engagement during play.

The individual differences (Figure 7A) in Coordinated JA reflect the joint contribution of infant and parent, and thus differ completely from measures obtained in discrete trial-based experiments that focus on just the infant's ability to follow but to not lead. Some infants and parents engaged in Coordinated JA for less than .10 of the play session while some did so for almost .40 of the play session. These are some of the relevant individual differences in Coordinated Joint Attention that represent everyday interactions and are predicted in Section III by the infants' performance on the discrete trial procedure measuring Joint Attention skills.

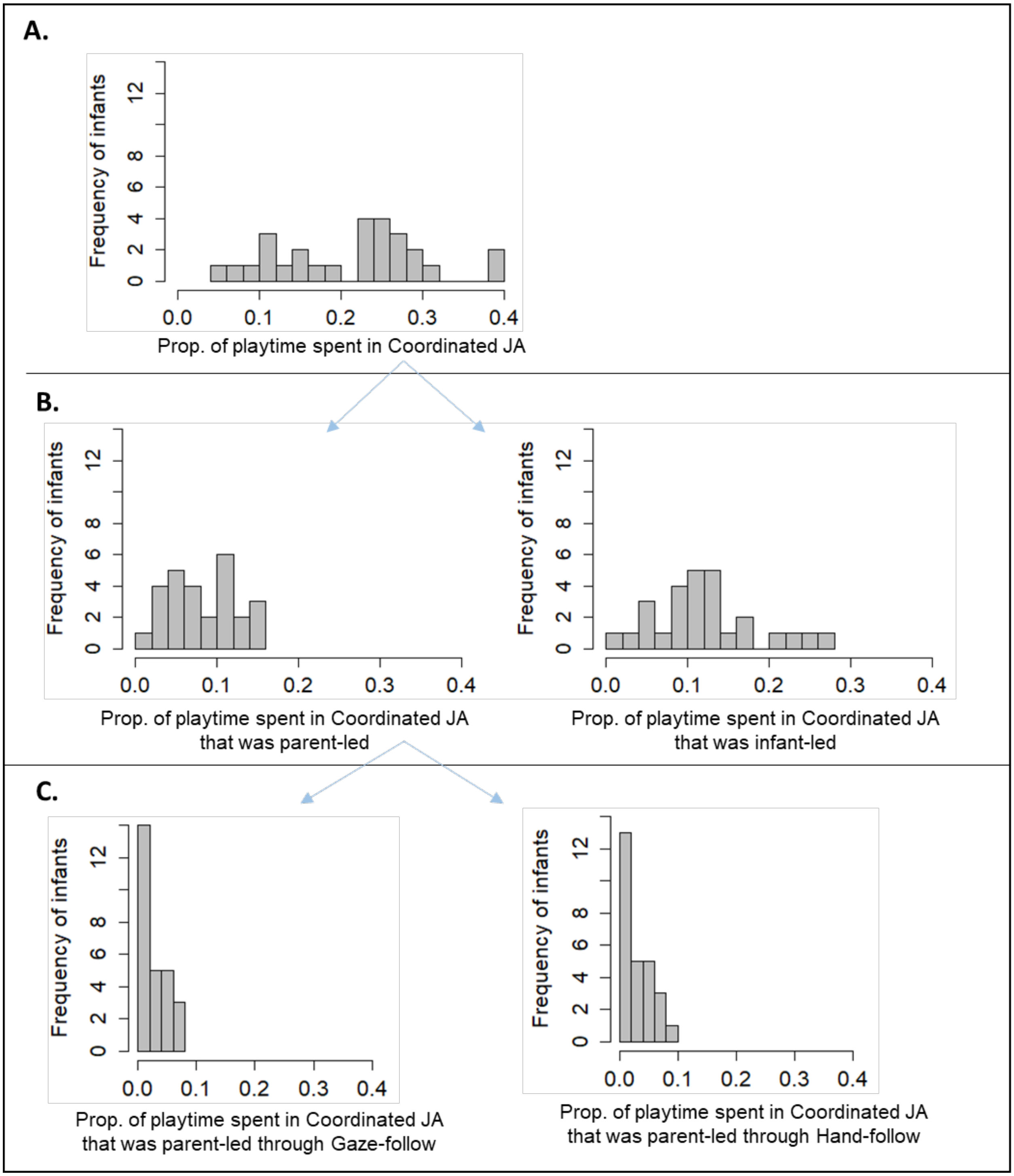


Figure 7. Individual differences in proportion of playtime spent in Coordinated JA (A), Coordinated JA that was parent-led or infant-led (B), and Coordinated JA that was parent-led and achieved through Gaze-following or Hand-Following (C). The arrows represent how the

different types of Coordinated JA moments are nested within each other starting with Coordinated JA (A), which includes all the other moments shown in B and C.

Individual differences in Coordinated JA parent-led and infant-led.

Consistent with previous research showing the dual role that infants play for the establishment of Coordinated JA in the context of play, we classified moments of Coordinated JA as Parent-led or Infant-led based on the partner who initiated the Coordinated JA bout by looking first to the object of the Coordinated JA moment. Individual differences in parent-led Coordinated Joint Attention captured the infant's ability to follow the parent's cues used when attending to an object while individual differences in infant-led Coordinated Joint Attention captured the parent's ability to follow the infant's cues. On average, 38% (SD=12, min=10, max=57) of all the Coordinated Joint Attention moments were initiated by the infant following the parent's look to an object (JA parent-led) while 62% (SD=12, min=43, max=90) were initiated by the parent following the infant's look to an object (JA infant-led).

Dyads engaged in "Coordinated JA parent-led" on average for 0.08 (SD=0.04) of the play session. The individual differences around this mean are shown in Figure 7B. Some infants engaged in parent-led Coordinated JA for 0.02 of the play session while some did so for 0.15. These specific type of Coordinated JA moments occurred on average 2.50 times per minute (SD=1.20) and lasted 1.78 seconds (SD=0.58) as Table 2 shows.

Dyads engaged in "Coordinated JA infant-led" on average for 0.12 (SD=0.06) of the play interaction with larger individual differences in this proportion of time spent as shown in Figure 7B. Coordinated Joint Attention bouts initiated by the infant occurred on average 3.84 times per minutes (SD=1.51) and lasted 1.71 seconds on average (SD=0.62). A non-parametric statistical test was used to compare the proportion of playtime spent in Parent-led versus Infant-led

Coordinated JA given the skew present in the differences scores for the subjects. A Wilcoxon signed-rank test revealed there is a significant difference with dyads spending more playtime in infant-led Coordinated JA as compared to parent-led Coordinated JA, v=71, p=0.003. This result is consistent with the idea that at this young age, Coordinated Joint Attention occurs more because parents tune their actions to the infant's and follow more than lead. These results support the Disconnect hypothesis because the "social interaction" recreated in the context of trial-based experiments lacks the most common type of joint attention that occurs during real-world interactions, the infant-led moments of coordinated visual attention.

Individual differences in pathways to Coordinated JA parent-led.

According to the Yu & Smith characterization of the pathways to engage in Coordinated JA, we examined whether Coordinated Joint Attention that was initiated by the parent occurred through different pathways, or cues that were potentially used by the infant to join the parent's look to an object. Adopting Yu & Smith (2017a) definitions, we classified all Parent-led Coordinated JA bouts as occurring through Gaze-following or Hand-following. The period right before the episode was analyzed to look for presence of a face look to the parent's face (to code for Gaze-follow) and when no face look was found, analyzed for the presence of parent's holding of the attended object during this window of time preceding the Coordinated JA moment (to code for Hand-follow). An individual infant, on average, had 2.85 (SD=2.27, min= 0.00, max=9.00) bouts of parent-led Coordinated JA through Gaze-following during the entire play session, and 3.44 (SD=2.85, min=0.00, max=11.00) that were parent-led Coordinated JA through Hand-following during the entire play session. Coordinated JA moments that were parent-led and occurred through Gaze-following and Hand-following accounted for 0.29 (SD=0.22, min=0.00, max=0.75) and 0.32 (SD=0.19, min=0.00, max=0.85), respectively, of all the

Coordinated JA moments that were parent-led. The remaining bouts of Coordinated JA (a third on average) did not occur through any of these pathways and are an open question to explore for future research.

Individual differences in infants' abilities to engage in parent-led Coordinated Joint attention through Gaze-follow or Hand-follow strategies are shown in Figure 7C. The proportion of playtime spent in parent-led Coordinated JA through Gaze-following was on average 0.03 (SD=0.02). This type of Coordinated JA bouts was extremely rare as it occurred 0.73 times per minute (SD=0.56) and lasted on average 1.81 seconds (SD=1.71). The proportion of playtime spent in parent-led Coordinated JA through Hand-following was on average 0.03 (SD=0.03). Coordinated Joint Attention bouts in which both the infant did not look at the parent's face and the parent made hand contact with the target of Coordinated JA, were also extremely rare as they occurred 0.91 times per minute (SD=0.76) and lasted on average 1.76 seconds (SD=1.17). These results are consistent with the Disconnect hypothesis because the "social interaction" recreated in the context of trial-based experiments lacks cues that lead to Coordinated JA as often as the gaze cues do in the context of real-world interactions.

The individual differences just reported on parent-led Coordinated Joint Attention occurring through Gaze- and Hand-following were a subset of all the Coordinated JA moments and they capture how often moments that could be due to only the infant following through different cues occurred. These estimates are thus based on only a few episodes but are still predictors relevant for the goal of this work, which as opposed to capturing differences between infants in their most common behavior (all Coordinated JA), they captured individual differences in how often these rare and very specific moments occur. Therefore, although all infants spent less than 0.10 of the play session in moments of parent-led Coordinated JA due to Gaze-following, there was some distribution with some infants spending less than 0.02 but some above that.

Together, the results from this section show that consistent with previous work, Coordinated JA occurred from the mutual contribution of parent and infant although for infants this age, parents are more likely to follow the infant-led Coordinated JA moments. Moreover, these results also show that there are alternative pathways that involve the hands of the parent to engage in Coordinated JA and infants differed too in the proportion of playtime spent in these exceptional moments of coordinated visual attention. These results are consistent with the Disconnect hypothesis since the behavior elicited in the context of free play differs from the phenomenon elicited in the context of discrete trial-based experiments.

Section III. Relations Between Discrete Trial-based Experiment and Continuous Free Play

Given all the individual differences in joint attention obtained in the discrete trial-based experiment and in the continuous free play session, are they related in any way? One possibility is that infant behavior across contexts is related because it is the same infant after all. Another possibility under the Disconnect hypothesis, is that the discrete trial-based experiment is sufficiently different, and disconnected, from free ranging behavior and thus elicits infant behavior, at all levels that does not play out in real life. Figure 8 shows the different constructs and infant behaviors that were tested to examine the Disconnect between discrete trial-based JA and the more naturalistic context of the free play session. The Disconnect is shown if: 1: JA is not related to Coordinated JA (Test 1), 2: JA is not related to infant behaviors that support word learning during play such as infant attention, infant sustained attention, infant attention to objects during naming and off-task behavior during play (Test 2), and if 3: at the very basic level the infant sensory-motor abilities that support JA and Coordinated JA (shown inside the grey ovals) are not related to each other (Test 3). All tests confirmed the Disconnect hypothesis and are presented next, one by one. Best practices in statistics advise researchers to limit the number of statistical tests performed in order to reduce Type I error, or the probability that the statistical test incorrectly rejected the Null hypothesis finding a significant relation where there was none. Here, even after inflating Type I in order to make an exhaustive examination of the Disconnect, no relations were found.

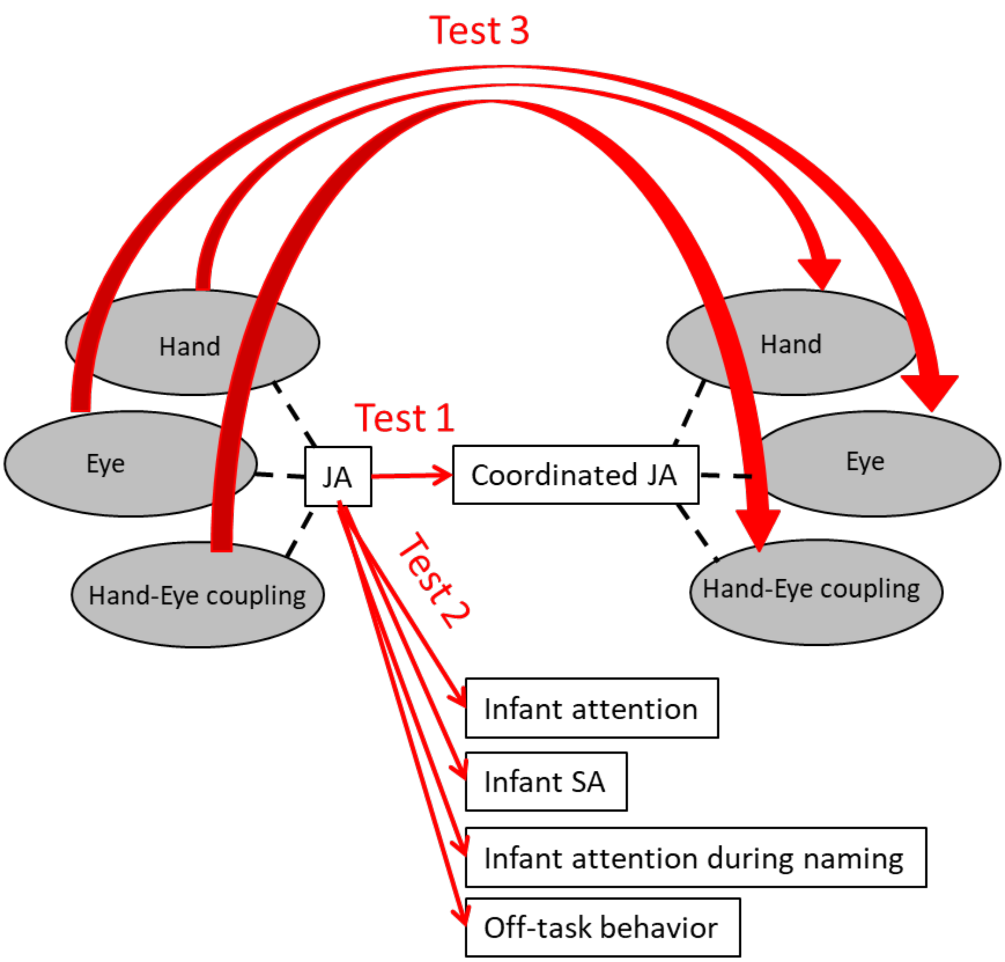


Figure 8. Visual representation of the tests carried out to test the Disconnect between the infant behaviors elicited by the discrete Trial-based experiment (left side) and by the continuous free play session (right side). The additional infant behaviors that occur during play and may support

learning as Coordinated JA does are shown too in the diagram below Coordinated JA. Finally the diagram shows in grey ovals the general sensory-motor behaviors that underlie infant behavior but in particular JA and Coordinated JA (theoretical links are shown in back dashed lines).

Test 1: JA does not predict Coordinated JA.

The central question of the study is the relation between measures of infant's joint attention abilities obtained during parent-infant play and during the discrete trial-based experiment (Test 1), phenomena that are referred to by the same name. Because the experimental studies are used to measure JA and act as a powerful diagnostic tool for language development, here we asked does performance in the traditional task (Gaze trials), and its variation implemented in the present study (Gaze-Hold trials), predict the time spent in *Coordinated* JA during parent-infant play?

Multiple linear regression was used to address the question, using the proportion of playtime spent in Coordinated JA (and its subtypes, Figure 7) during play as the dependent variable while the independent variables were the chosen first-look (Measure 4) and all-trial (Measure 1) measure of JA ability during the discrete trial-based task. A total of six models were tested as part of Test 1 shown in Figure 8, they differed in the dependent variable of Coordinated JA during play that was being predicted (Coordinated JA (x2), Coordinated JA parent-led (x2), Coordinated JA parent-led through Gaze-Follow and Coordinated JA parent-led through Hand-Follow); three models used measures from the Gaze trials as the independent variables while the other three models used measures from the Gaze-Hold trials, which provided an estimate of infant's JA abilities in a context more similar to the real world. All models met the assumptions of multiple linear regression including random and normally distributed residuals, no multicollinearity and constant variance. The dependent variables in Models 5 and 6, the proportion of playtime spent in parent-led Coordinated JA through Gaze-following and through Hand-following, respectively, were log-transformed to meet the assumptions of the models. Each model was compared to a Null model, which shared the same dependent variable but only estimated an intercept and no other predictors, using two widely used criteria for model selection that balance the level of fit of the model with the complexity of the model represented by the number of independent variables used (Akaike information criterion and the Bayesian information criterion).

In Model 1, the proportion of playtime spent in Coordinated Joint Attention for an individual dyad was predicted by the infant's mean proportion of the Gaze-cue time spent in JA (measure shown in Figure 6B) and by the infant's proportion of correct Gaze trials (measure shown in Figure 6D). The multiple regression was not significant ($F(2,24)=0.34$, n.s), with an adjusted R^2 of -0.05. Table 3 shows the coefficients of the model and associated t statistics and p-values. Model 1 was compared to a null model, which had no predictors but only used the intercept as the predictor, using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Burnham & Anderson, 2004). The results suggest the Null model was preferred over Model 1 ($AIC_{Null}=-50.07$, $AIC_{Model1}=-46.84$; $BIC_{Null}=-47.48$, $BIC_{Model1}=-41.66$).

Table 3

Model 1: Predicting Coordinated JA from JA in Gaze trials

	B	**SE**	**t**	**p**
Intercept	0.19	0.08	2.39	0.02
Mean prop. of cue in JA during Gaze trials	-0.01	0.16	-0.08	0.94
Prop. of correct Gaze trials	0.06	0.07	0.83	0.42

In Model 2, the proportion of playtime spent in Coordinated Joint Attention for an individual dyad was predicted by the infant's mean proportion of the Gaze-Hold cue time spent in JA (measure shown in Figure 6A) and by the infant's proportion of correct Gaze-Hold trials (measure shown in Figure 6C). Even when trial-base experiments included the cues that are available for infants to follow the adult's engagement with an object in real-world contexts, infant behavior in the experiment did not predict the degree to which an infant can engage in Coordinated JA during free play. The multiple regression was not significant ($F(2,24)=0.30$, n.s), with an adjusted R^2 of -0.05. Table 4 shows the coefficients of the model and associated t statistics and p-values. Model 2 was compared to a null model, which had no predictors but only used the intercept as the predictor, using the AIC and BIC. The results suggest the Null model was preferred over Model 2 (AIC_{Null}=-50.08, AIC_{Model2}=-46.75; BIC_{Null}=-47.48, BIC_{Model2}=-41.57). Together, the results of Model 1 and Model 2 confirmed the Disconnect hypothesis since individual differences in Coordinated Joint Attention during continuous free play were not related to individual differences in Joint Attention during the discrete trial-based tasks. We turn to examine the predictive value of measures obtained in discrete trials for Coordinated Attention that was parent-led.

Table 4

Model 2: Predicting Coordinated JA from JA in Gaze-Hold trials

	B	**SE**	**t**	**p**
Intercept	0.34	0.17	2.03	0.05
Mean prop. of cue in JA during Gaze-Hold trials	-0.14	0.20	-0.68	0.50
Prop. of correct Gaze trials	-0.02	0.11	-0.20	0.85

Notes. R^2= 0.02 ($p>0.05$)

In Model 3, the proportion of playtime spent in Coordinated Joint Attention that occurred because the infant followed the parent's look (parent-led coordinated JA) was predicted by the mean proportion of the Gaze-cue time spent in JA (measure shown in Figure 6B) and by the proportion of correct Gaze trials (measure shown in Figure 6D) for each subject. The multiple regression was not significant ($F(2,24)=1.26$, n.s), with an adjusted R^2 of 0.02. Table 5 shows the coefficients of the model and associated t statistics and p-values. Model 3 was compared to a null model, which had no predictors but only used the intercept as the predictor, using the AIC and BIC. The results suggest the Null model was preferred over Model 3 ($AIC_{Null}=-93.40$, $AIC_{Model3}=-92.10$; $BIC_{Null}=-90.81$, $BIC_{Model3}=-86.91$).

Table 5

Model 3: Predicting Coordinated Parent-led JA from JA in Gaze trials

	B	**SE**	**t**	**p**
Intercept	0.06	0.03	1.86	0.07
Mean prop. of cue in JA during Gaze trials	-0.02	0.07	-0.24	0.81
Prop. of correct Gaze trials	0.05	0.03	1.56	0.13

Notes. R^2= 0.09 ($p>0.05$)

In Model 4, the proportion of playtime spent by each infant in parent-led Coordinated Joint Attention was predicted by the infant's mean proportion of the Gaze-Hold cue time spent in JA (measure shown in Figure 6A) and by the infant's proportion of correct Gaze-Hold trials (measure shown in Figure 6C). The multiple regression was not significant ($F(2,24)=0.44$, n.s), with an adjusted R^2 of -0.04. Table 6 shows the coefficients of the model and associated t statistics and p-values. Model 4 was compared to a null model, which had no predictors but only used the intercept as the predictor, using the AIC and BIC. The results suggest the Null model was preferred over Model 4 ($AIC_{Null}=-93.40$, $AIC_{Model4}=-90.37$; $BIC_{Null}=-90.81$, $BIC_{Model4}=-$

85.19) and thus suggest that, as hypothesized, individual differences in neither Gaze-trials nor in Gaze-Hold trials was related to individual differences in the degree to which infants engaged in Parent-led Coordinated JA during play.

Table 6

Model 4: Predicting Coordinated Parent-led JA from JA in Gaze-Hold trials

	B	**SE**	**t**	**p**
Intercept	0.12	0.07	1.57	0.13
Mean prop. of cue in JA during Gaze-Hold trials	-0.07	0.09	-0.78	0.44
Prop. of correct Gaze-Hold trials	0.03	0.05	0.71	0.49

Notes. R^2= 0.03 ($p>0.05$)

The predictors used for in Models 5 and 6 were obtained from infant's performance in Gaze and Gaze-Hold trials, respectively, because conceptually Gaze trials, and not Gaze-Hold trials, mirror more closely the type of Coordinated JA that was parent-led through Gaze-following. Gaze-Hold trials mirror more closely the type of Coordinated JA that was parent-led through Hand-following because they also include a holding cue that tested infant's abilities to pick up hold cues.

In Model 5, the proportion of playtime spent by each infant in parent-led Coordinated Joint Attention occurring through Gaze-following was predicted by the infant's mean proportion of the Gaze cue time spent in JA (measure shown in Figure 6B) and by the infant's proportion of correct Gaze trials (measure shown in Figure 6D). The multiple regression was not significant ($F(2,24)=1.93$, n.s), with an adjusted R^2 of 0.07. Table 7 shows the coefficients of the model and associated t statistics and p-values. Model 5 was compared to a null model, which had no predictors but only used the intercept as the predictor, using the AIC and BIC. The results suggest the Null model was preferred over Model 5 (AIC_{Null}=-53.04, AIC_{Model5}=-53.06; BIC_{Null}=-50.45, BIC_{Model5}=-47.88). Together, these results show that individual differences in discrete trial-based tasks that use gaze cues are not predictive of even the most closely matched type of Coordinated JA occurring through Gaze-following during continuous free play session.

Table 7

Model 5: Predicting Coordinated Parent-led JA that occurred through Gaze-following during play from JA in Gaze trials

	B	**SE**	**t**	**p**
Intercept	0.03	0.07	0.44	0.66
Mean prop. of cue in JA during Gaze trials	0.11	0.14	0.76	0.45
Prop. of correct Gaze trials	0.12	0.07	1.82	0.08

Notes. R^2= 0.14 (p>0.05). Dependent variable was log-transformed.

In Model 6, the proportion of playtime spent by each infant in parent-led Coordinated Joint Attention occurring through Hand-following was predicted by the infant's mean proportion of the Gaze-Hold cue time spent in JA (measure shown in Figure 6A) and by the infant's proportion of correct Gaze-Hold trials (measure shown in Figure 6C). The multiple regression was not significant ($F(2,24)=0.20$, n.s), with an adjusted R^2 of -0.07. Table 8 shows the coefficients of the model and associated t statistics and p-values. Model 6 was compared to a null model, which had no predictors but only used the intercept as the predictor, using the AIC and BIC. The results suggest the Null model was preferred over Model 6 (AIC_{Null}=-52.97, AIC_{Model6}=-49.41; BIC_{Null}=-50.37, BIC_{Model6}=-44.22). The results suggest that individual differences in discrete trial-based tasks that use gaze plus hold cues are not predictive of even the most closely matched type of Coordinated JA occurring through Hand-following during free continuous play.

Table 8

Model 6: Predicting Coordinated Parent-led JA that occurred through Hand-following during play from JA in Gaze-Hold trials

	B	**SE**	**t**	**p**
Intercept	0.10	0.16	0.63	0.54
Mean prop. of cue in JA during Gaze-Hold trials	0.01	0.20	0.05	0.96
Prop. of correct Gaze-Hold trials	0.06	0.10	0.59	0.56

Notes. R^2= 0.02 (p>0.05). Dependent variable was log-transformed.

All the models reported in this section predicted Coordinated JA from the two relevant first-look and all-trial based measures of the discrete trial-based experiment (prop. of cue time spent in JA and prop. of correct trials). The Supplemental Material reports six additional models that predict real-world Coordinated JA from a larger set of variables obtained from the discrete trial-based experiment, such as the latency to look at the cued object by the experimenter or the proportion of the cue time spent looking at the experimenter's face. The results from these additional models further support the Disconnect hypothesis because even after leveraging on the use of six to eight variables to capture as much as possible of the infants' performance during the experiment, the results show that individuals' performance in the discrete trials as described by these many other measures together are not predictive of the infant's proportion of playtime spent in Coordinated JA, Coordinated parent-led JA, Coordinated parent-led JA through Gaze- and Hand-following. Together all the results suggest, as predicted, that the phenomenon elicited in the discrete trial-based experiment has nothing to do with the phenomenon elicited in the real world.

Test 2: JA and infant behaviors that support learning.

A strong test of the Disconnect hypothesis is to examine the relation between JA and infant behaviors occurring during free play that support word learning, just as JA supposedly predicts word learning as well (Test 2). The JA ability demonstrated in the Discrete Trial-based experiment is supposedly associated, or claimed to bring about, these infant behaviors (shown in Figure 8) because an increased ability to follow cues when interacting with a social partner helps to build an optimal joint action around the objects that predicts learning and can directly influence, or at least be related to, these infant behaviors that occur during play. Under this view, infants who are more sensitive to gaze and/or gaze-hold cues, are infants who show more attention to objects, more sustained attention, more attention to named objects during naming and less off-task behavior. The following analyses were used to carry out Test 2 for the Disconnect hypothesis.

First, we tested whether JA was related to the amount of time that infants' spent looking to objects during the play session with the parent. Infant attention to objects was measured as the proportion of playtime spent looking at objects. Figure 9A shows subject histograms of this measure since infants ranged from looking at objects for 0.28 up to 0.86 of the play session (M=0.56, SD=0.15).

In the same way, we tested whether JA was related to the infant's ability to attend to individual objects for several seconds during play. We measured infant sustained attention (SA) during the free play session as the proportion of playtime spent in sustained attention (defined as unbroken visual attention to an object that lasts 3 seconds or longer) by infants. Figure 9B shows individual differences with respect to this infant behavior since infants engaged in SA for as low as 0.06 of the interaction for as high as 0.66 of the interaction (M=0.32, SD=0.15).

Third, we tested whether JA was related to the infant attention during all the naming utterances that the parent produced and computed the proportion of time during the naming utterance in which the infant looked to the named object, or target of the naming utterance. All the naming utterances and the proportion of the naming utterance spent in looking at the target by the infant was averaged for each infant to produce a distribution of individual differences shown in Figure 9C. Infants ranged from 0.19 to 0.64, with some infants being on average very good at attending to the named object for almost half of the naming utterance (M=0.41, SD=0.11).

Finally, we tested whether JA was related to infant's interest and willingness of play with the parent and objects. We measured infants' off-task behavior as the proportion of playtime spent in looking at other objects by the infant around the room. Figure 9D shows subject histograms of this measure as infants were off-task for as low as 0.08 of the interaction and for as high as 0.56 of the interaction (M=0.24, SD=0.11).

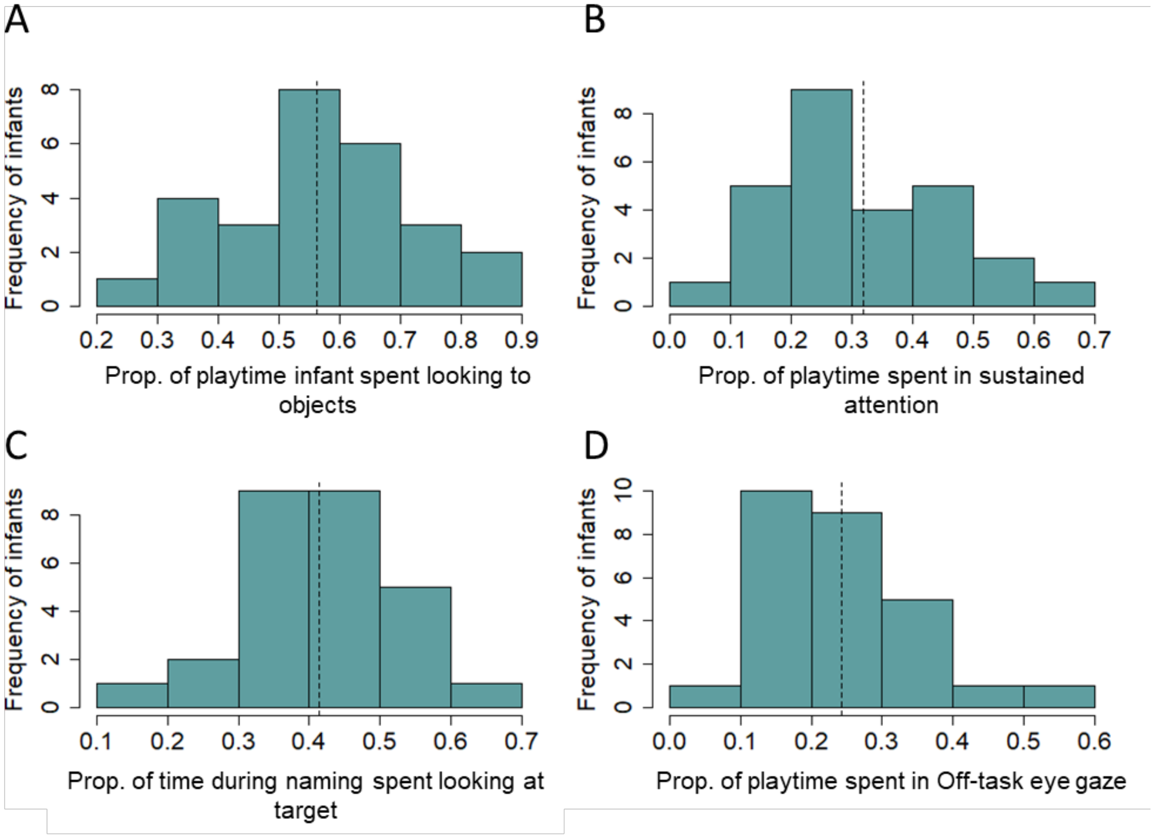


Figure 9. Subject histograms showing individual differences in infant behaviors that relate to learning during play. Dotted black lines indicate the mean of each frequency distribution.

Are these individual differences related to each other? The answer appears to be yes. Figure 10 shows both the scatterplots and Pearson correlation coefficients describing the association between these infant behaviors that relate to learning during free play. Infants who tended to spend a large proportion of playtime looking at objects and sustaining attention, were those who looked to the object named by the parent more and those who spent less time in off-task behavior.

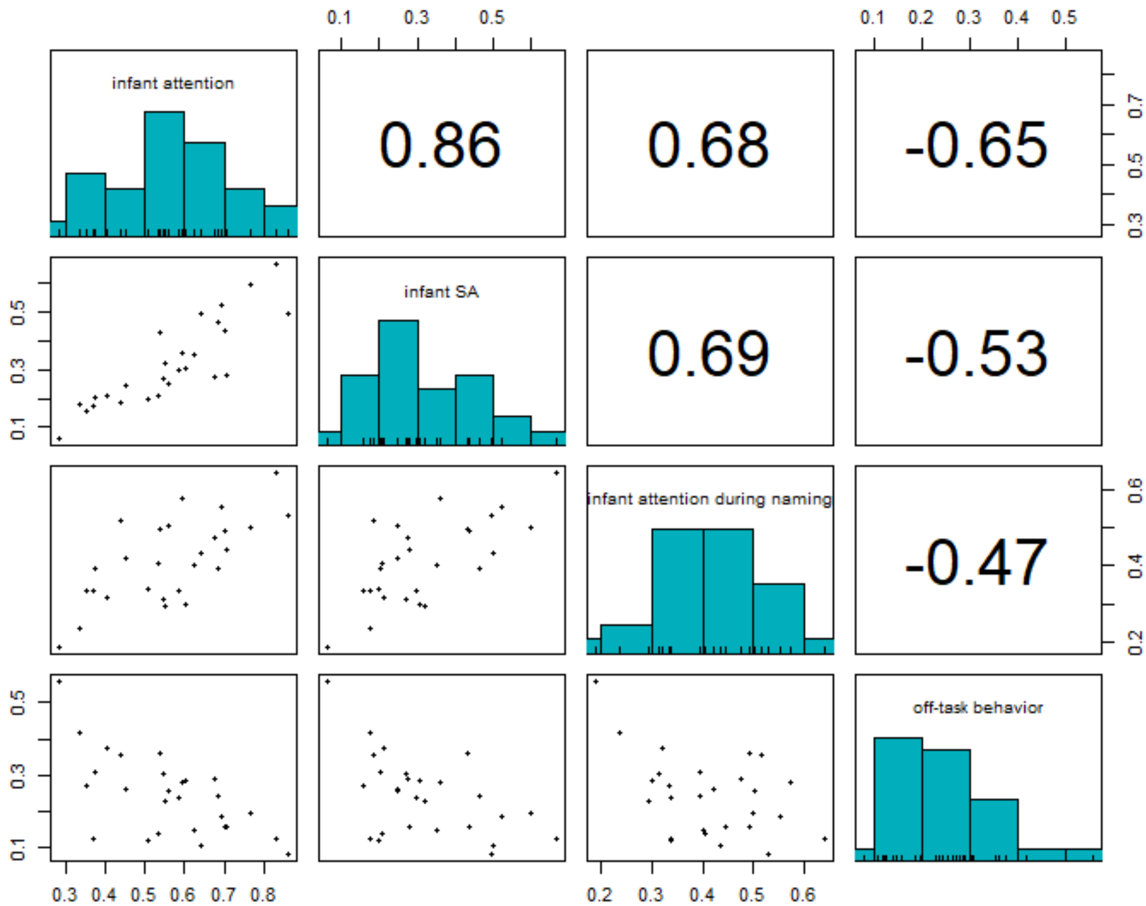


Figure 10. Histograms of infant behaviors that support learning and scatterplots between them. Because all measures are normally distributed, the figure also shows the Pearson correlation coefficients between each pair of measures.

The relevant question is whether the infant behaviors that support learning during play are related to JA during the discrete trial-based experiment. We ran eight multiple linear regression models. Four models predicted individual differences in infant behaviors during play from the two chosen all-trial and first-look measures in Gaze trials; four additional models predicted the same infant behaviors during play but used performance in Gaze-Hold trials as the independent variables of the models.

Table 9 identifies each model, the dependent variable predicted by each model as well as independent variables used in each model. Moreover, it provides the F test of the model's fit as well as the associated p-value and adjusted R square. The individual estimates of the model coefficients were not included since they were all non-significant. Each model implemented was compared to a Null model that did not have the independent variables from the trial-based experiment, and thus the last column lists the model that had the lowest AIC and BIC values and was thus preferred. The results show that none of the infant behaviors considered are predicted by infants' performance in neither Gaze nor Gaze-Hold trials. In fact, for the case of every model, the Null model provided a better fit of the data.

Table 9

Summary of results obtained in Models 7-14.

Model	Dependent Variable	Independent Variables	$F(2,24)$	p	Adjusted R^2	Model preferred
7	Prop. of playtime the Infant spent looking at objects	Gaze trials	0.14	0.87	-0.07	Null
8	Prop. of playtime spent in Infant Sustained Attention	Gaze trials	0.02	0.98	-0.08	Null
9	Mean prop. of time inside the naming utterance spent looking at target	Gaze trials	0.08	0.92	-0.07	Null
10	Prop. of playtime spent in Infant off-task	Gaze trials	0.12	0.88	-0.07	Null
11	Prop. of playtime the Infant spent looking at objects	Gaze-Hold trials	0.81	0.45	-0.01	Null
12	Prop. of playtime spent in Infant Sustained Attention	Gaze-Hold trials	0.35	0.71	-0.05	Null
13	Mean prop. of time inside the naming utterance spent looking at target	Gaze-Hold trials	0.17	0.85	-0.07	Null

14	Prop. of playtime spent in Infant off-task	Gaze-Hold trials	0.57	0.57	-0.03	Null

The results presented in this section (Test 2) are consistent with the Disconnect hypothesis as they showed there is no relation between measures of JA obtained in discrete trial-based experiments and measures of infant behaviors during play that are conducive to learning including infant attention to objects, infant sustained attention, infant attention to the named object during an utterance and infant off-task looking behavior during play. These eight additional models provide strong evidence that there is a Joint Attention Disconnect in which the individual differences picked up by the traditional gaze-follow and the implemented gaze-and-hand follow tasks, are not predictive of meaningful individual differences occurring in the continuous free play session. These results suggest that gaze-following itself may not contribute to infant's everyday behaviors that support learning and therefore they raise the possibility that gaze-following may be only a marker that predicts language development but not the actual driver and cause of language development.

Test 3: Sensory-motor behaviors across contexts.

The behavior elicited by trial-based experiments of JA is a good diagnostic tool that predicts optimal development. One open question is what are the underlying sensory-motor skills supporting this behavior in discrete trial-based experiments? Are those skills the same skills that support infant behavior in free play or not? Does the Disconnect apply to even low-level more general behaviors that support JA as well as Coordinated JA? The analyses that constitute Test 3 investigate whether individual differences in the infant's general sensory-motor skills that were recruited in the Trial-based Task were related to individual differences in the same skills used in the everyday Play Interaction with the parent. We ask, for instance, were infants who attended more to objects in discrete trials those who attended more to objects during free play? There are two possibilities. On the one hand, we would expect that sensory-motor infant abilities would be correlated across task because it is the same system acting on objects with social partners but doing that in two different contexts. On the other hand, under the Joint Attention Disconnect hypothesis, it is possible that even though the same infant is tested in each context there are perturbations in the discrete trials that break the coupling inherent in the infant's system that produces the abilities and thus the sensory-motor abilities recruited in discrete trials are not related to the abilities used during every day free play with the parent. Accordingly, we would predict that infants who looked more to objects during the discrete trial-based task are not the same infants who looked more to objects during play with the parent. This possibility could suggest that the behavior that is diagnostic of optimal development is not the "everyday" behavior of the system but the behavior of the system once perturbations of everyday dynamics are introduced.

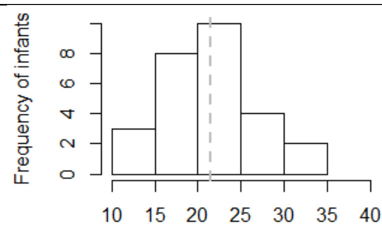
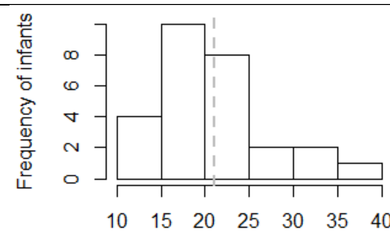
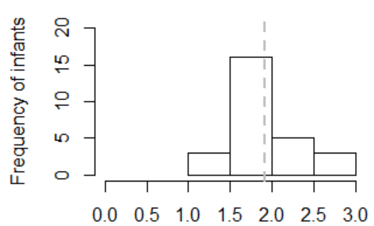
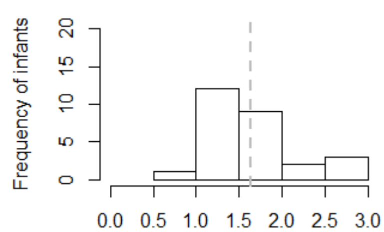
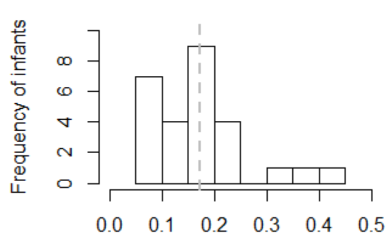
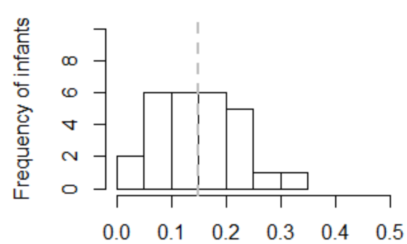
In order to address this question we examined the discrete trial-based experiment as a sensory-motor task, in which infants were actually asked to interact with an adult, and toys on a tabletop. We computed the same measures across both tasks effectively capturing the infants' attention to objects, sustained attention to objects, attention to faces, holding of an object and holding objects while attending visually to an object. We report the pairwise correlations to examine these "apples-to-apples" relations. For these analyses of infant's sensory-motor abilities both Gaze and Gaze-Hold trials were combined in order to estimate infant behaviors from the entire discrete trial-based experiment.

The first column of Table 10 lists the different measures obtained across contexts to measure the infants' general sensory-motor abilities. Each of the measures was plotted on the x-axis of the subject histograms shown in columns two and three of Table 10. The correlation coefficient shown in the fourth column of the table indicates the strength of the association between these two measures for an infant in the sample. As it is clear, all the correlations are weak and not significant except for the correlation between proportion of time in which the infant looked to the adult's face, which was moderate but only marginally significant. Infants who spent more time during the trial-based task looking at the experimenter tended to be those who spent more playtime looking at the parent's face during play. Perhaps the preference to attend to face over objects and hands (Jayaraman, Fausey & Smith, 2017), which develops around this age, determines behavior in both contexts. This finding needs to be interpreted in the context it occurred because it is the only significant result out of the many statistical tests conducted here and thus, it may not be a real and replicable result outside this context. The overall pattern of results supports the Disconnect and in fact suggests that the behavior elicited by the Discrete Trial-based Experiment is *task determined* and therefore highly dependent on the context and the arbitrary decisions that researchers make to define the research context.

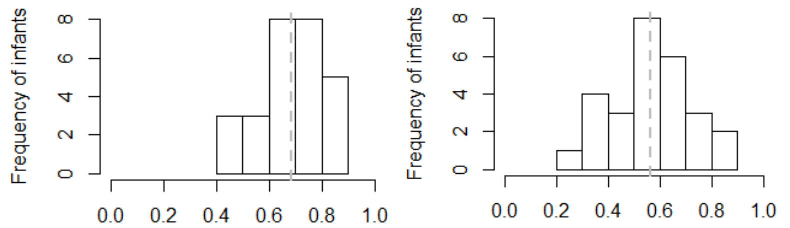
Interestingly, while the individual differences shown across tasks were not correlated, the overall distributions of measures on the infants' allocation of attention to objects and faces looked very similar to each other. For instance, the distributions of the frequency per minute of infant looks to objects in the discrete trial-based Task and in Free Play, were very similar and they were both centered on an almost identical mean. There were only relatively minor increases found in the Trial-based task compared to free play with respect to the duration of infant looks to objects, proportion of playtime looking at objects, frequency of looks to faces and overall proportion of playtime in looking at face. This is surprising given the very different contexts that infants faced in the study. These results may suggest that the different demands placed by the different tasks on the infant's visual attention *can* lead to similar and related patterns of attention for an individual across task but they don't. Discrete trial-based task may not impose a completely different set of demands on attention for a 9-month-old but it introduces perturbations to the infants' own system at the individual level and different infants respond differently in a pattern that cannot be generalized for all infants. That is, some infants look more to objects during the trial-based task compared to play, but some looked the same or less.

Table 10

Individual Differences Along Same Measures Across Task and Correlations Between Them

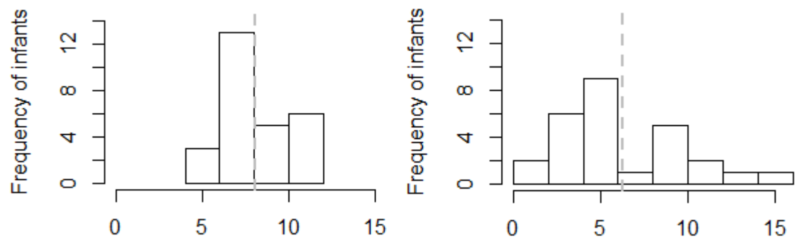
Measure	Trial-based Task	Continuous Free Play	Correlation
Frequency per minute of infant looks to objects			r_spearman= 0.30, *n.s*
Mean duration of infant looks to objects			r_spearman= 0.08, *n.s*
Proportion of infant looks to objects that were at least 3-seconds-long			r_spearman= 0.14, *n.s*

Proportion of time in which infant looked to objects



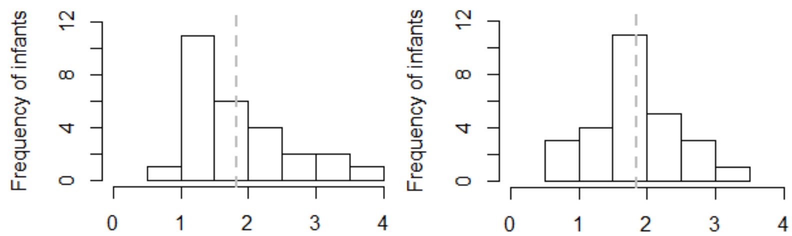
$r_{pearson}$= 0.04, *n.s*

Frequency per minute of infant looks to adult's face



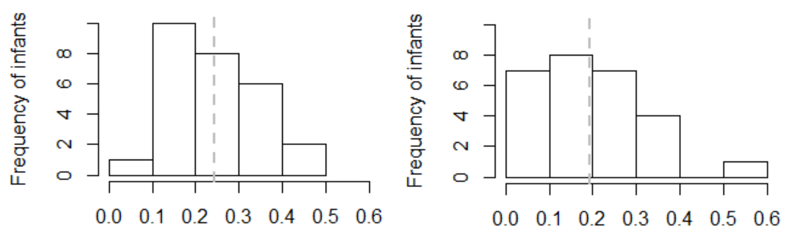
$r_{pearson}$= 0.05, *n.s*

Mean duration of infant looks to adult's face



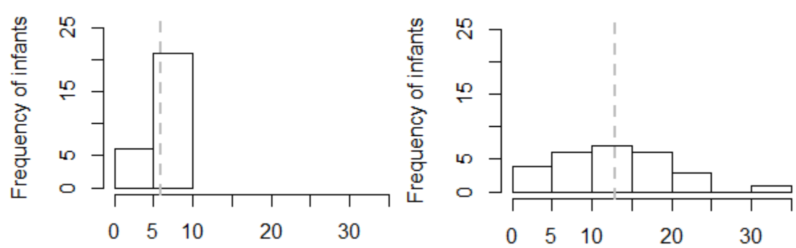
$r_{spearman}$= 0.07, *n.s*

Proportion of time in which infant looked to adult's face

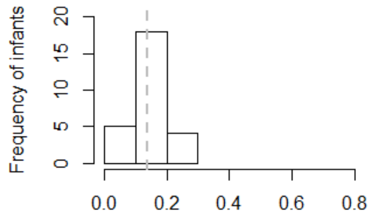
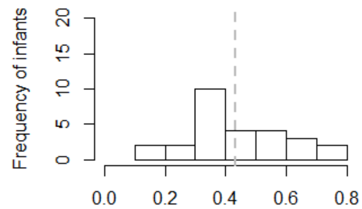


$r_{pearson}$= 0.38, *0.05^*

Frequency per minute of infant hand contact with an object



$r_{spearman}$= 0.19, *n.s*

Proportion of time inside infant look in which infant held the attended-object			r_spearman= 0.24, *n.s*

Notes. ^=p<.10; *=p<0.05. X-axis of histograms in shown in the "Measure" column of each row. Dotted line in each subject histogram indicates the mean of each distribution.

These data demonstrate that the discrete trial-based context for JA not only perturbed the infants' own dynamics of infant visual attention but also the infants' holding behaviors and the infants' coordination of eyes and hands. For most of the infants, hand contact with an object occurred more frequently in the context of free play but the infants who held more objects during play were not the same infants who held more objects during the trial-based task. This was also true for the likelihood of an infant to manually hold the objects they were visually attending to, which was higher overall during free play but unrelated to the likelihood to hold while looking for an infant during the discrete trial-based task. The discrete trial-based experiment changed infants' holding and hand-eye coordination.

Together, the results support the Disconnect hypothesis showing that even general low-level sensory-motor behaviors elicited by the discrete trial-based context are not related to sensory-motor behaviors used during play. The evidence suggests that discrete trials perturbs the system and breaks the coupling in the infants' system by changing for each infant how the eyes and hands behave and interact with each other. It is possible the behavior captured by the trial-based tasks is not the "everyday" behavior of the system but the response of each infant to those perturbations.

Section IV. Temporal Dynamics of JA and Coordinated JA

We further uncover the Joint Attention Disconnect phenomenon between widely used contexts for joint attention, by examining of the temporal dynamics of the phenomena elicited across context: the latencies between the onset of the infant's look and the adult's look.

We identified all the moments of Joint Attention (for the trial-based context) and moments of Coordinated Joint Attention (for the play context). Then for each of these, we subtracted the onset of the infant's first look to the adult-attended-object *from* the onset of the adult's look to the object. As a result, negative values indicate the number of seconds that it took for the infant to join the adult's cue/look; positive values indicate the number of seconds that it took for the adult to join the infant's look. Clearly the context of the discrete trial-based task took only negative values since it was not possible for the infant to begin a joint attention episode.

In the discrete trial-based experiment, there were 316 cues provided by the experimenter. Out of those, the infant joined 309 with a look to the cued object by the experimenter at least once at some point during the entire cue. On average, infants took -0.48 seconds to look at the object, with latencies ranging from -7.85sec to 0.00sec (SD=0.96), and the median lag was 0.00. Figure 11 shows the entire distribution of the latencies for infants and the experimenter in the discrete-trial task. An overwhelming majority of events fell within the bin of 0 and -0.1 seconds, which indicates that infants were extremely fast in following the cue.

12

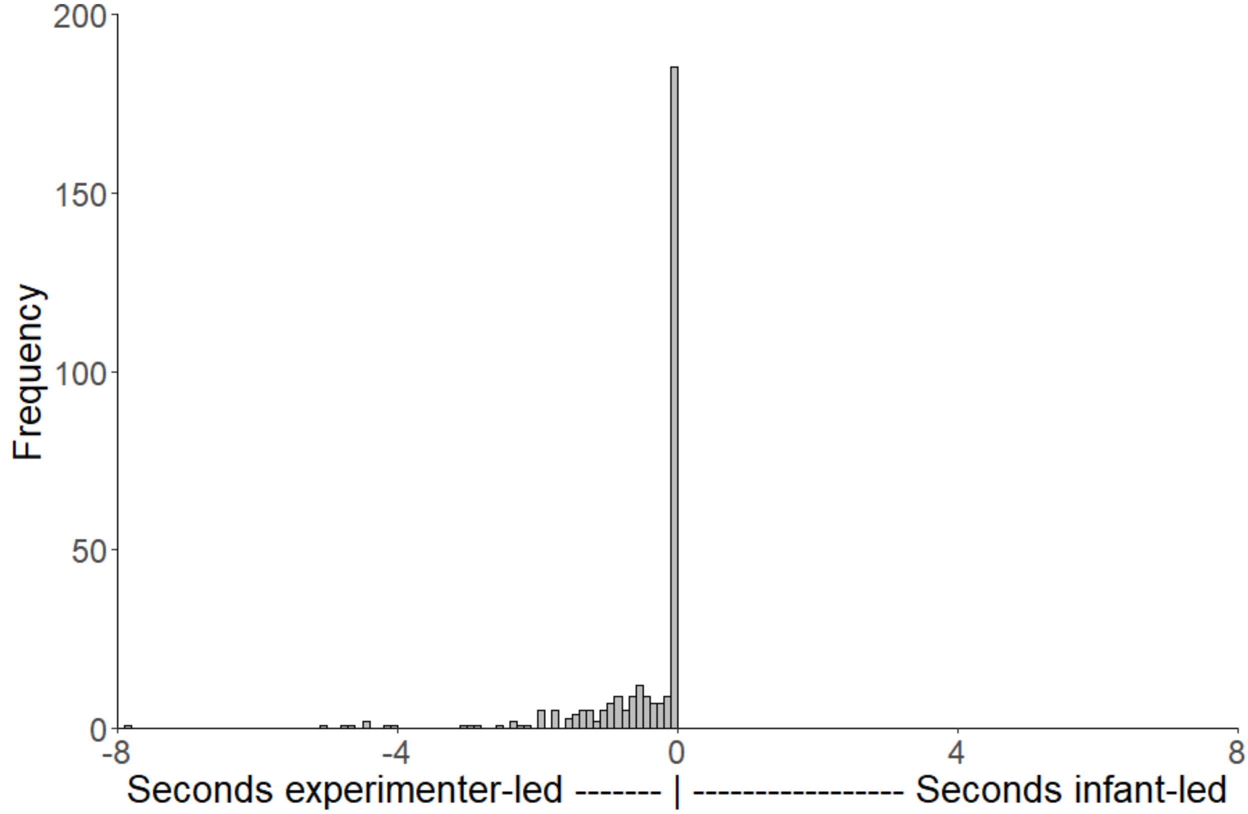
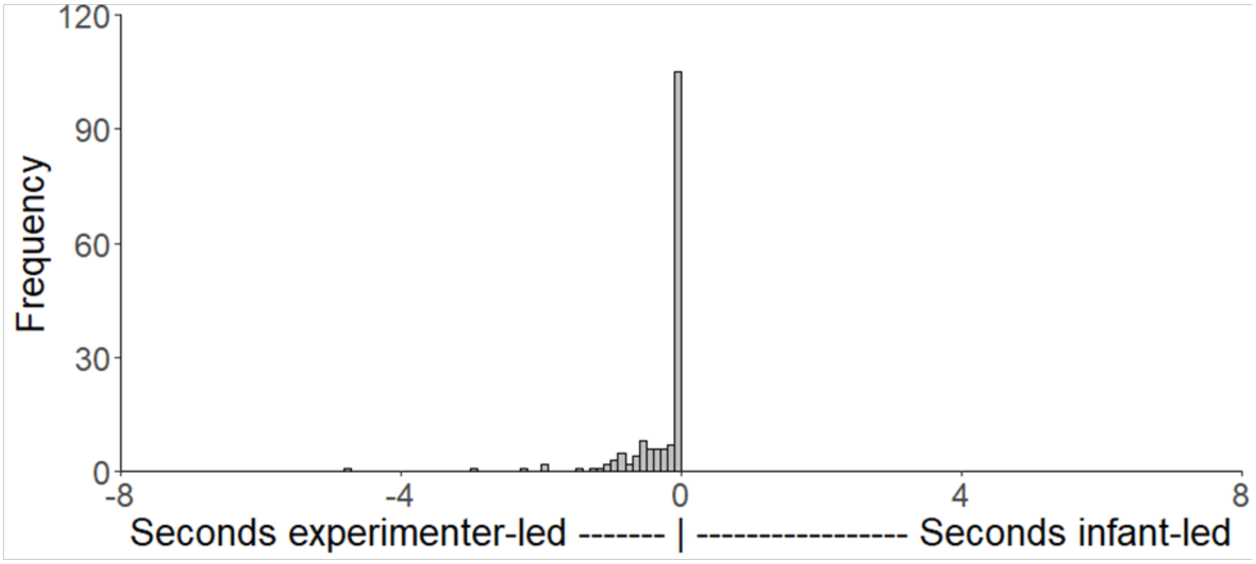


Figure 11. Latency between onset of experimenter's cue and onset of infant's first look to the cued-object.

Latencies were slightly different for Gaze and Gaze-Hold cues. Infants responded to Gaze-Hold cues slightly faster than to Gaze cues. The mean latency of infant joining a Gaze-Hold cue was -0.26 (-4.73 to 0.00, SD=0.58, median=0.00) while it was -0.72 (-7.85 to 0.00, SD=1.22, median=0.00) for Gaze cues. Figure 12 shows the entire distribution of latencies to Gaze and Gaze-Hold cues. Although Gaze cues have a longer and slightly heavier tail of latencies, both distributions look very similar: they both peak at the bin of latencies between 0 and 0.1 seconds, which is extremely fast.

A. Latency to JA in Gaze-Hold trials



B. Latency to JA in Gaze trials

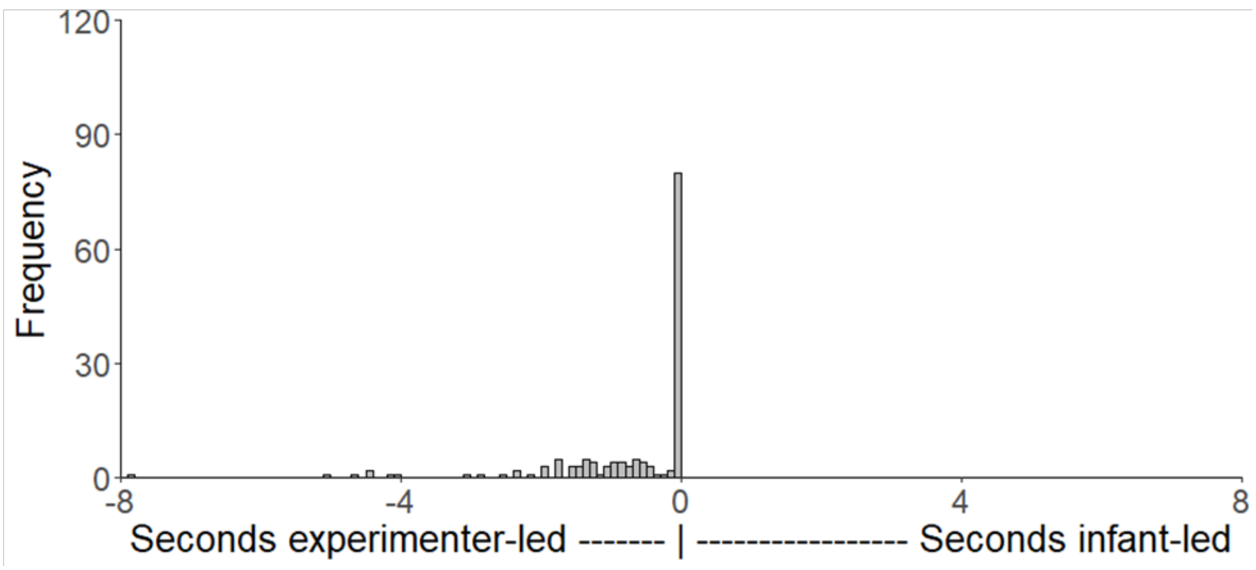


Figure 12. Latency between onset of experimenter's cue and onset of infant's first look to the cued-object separate for Gaze-Hold trials (A) and Gaze trials (B).

The latencies between infant and parent behavior for Coordinated Joint Attention bouts were the windows of time from the onset of the leader's look to the onset of the follower's look (or the onset of JA). That is, for parent-led Coordinated JA bouts the window started at the onset of the parent's look and ended at the onset of the parent-led JA bout; conversely, for infant-led Coordinated JA bout the window started at the onset of the infant's look and ended at the onset of the infant-led JA bout. There were a total of 672 episodes of Coordinated JA considered for this analysis. Figure 13 shows the entire distribution of the latencies to engage in these 672 Coordinated JA bouts between all the parent-infant dyads in the sample.

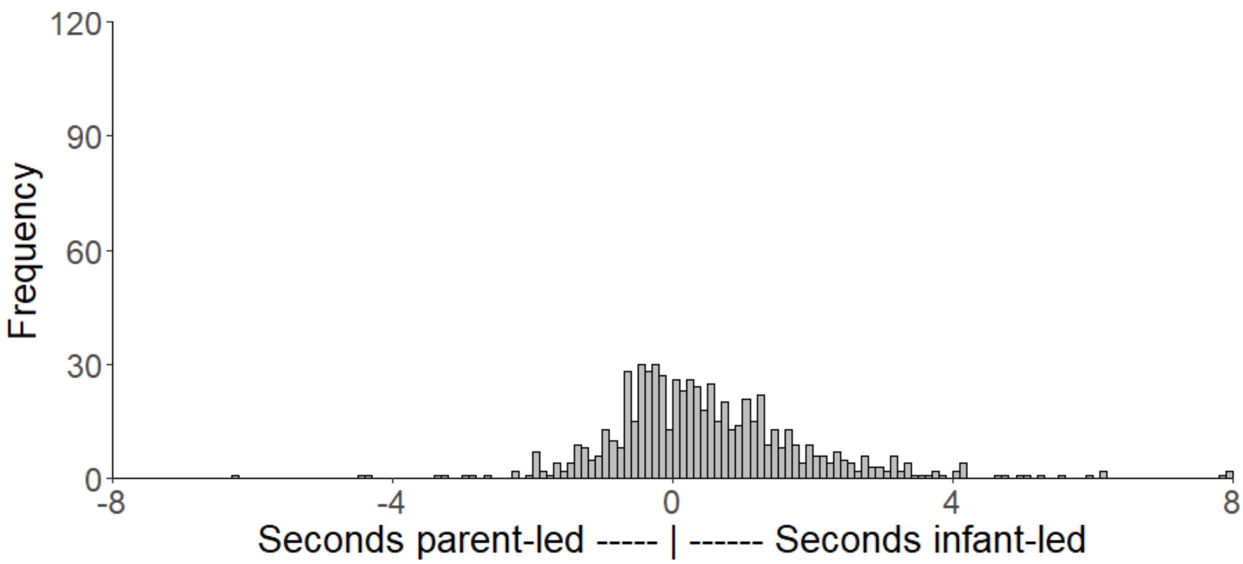


Figure 13. Latencies between onset of parent's first look and onset of infant's first look to the jointly-attended-object during continuous free play.

Figure 13 shows that the temporal dynamics of the Coordinated Joint Attention occurring during continuous free play are much different from the temporal dynamics of Joint Attention occurring in the discrete trial-based task. The distribution of Coordinated JA is not completely symmetric around the 0 point, which would represent instantaneous coordination of parent and infant eye-gaze, but has more events on the positive side. This suggests that infant-leading JA was more common for infants at this age and their parents when they engaged in play. In fact, the mean latency to engage in Coordinated Joint Attention was 0.53 seconds, which means that on average parents were more likely to join the infant's look. The latencies ranged from taking the infant -6.23 sec to joining the parent, to taking the parent a maximum of 12.36 seconds of parent to join the infant (SD=1.53sec, median=0.31). Nonetheless, most latencies to Coordinated JA were very short with 0.61 of them falling between -1 and 1 second, and 0.86 of them falling between -2 to 2 seconds. Although these data show there is tight coordination between infant and parent's eye-gaze when getting into Coordinated JA, the latencies are not as low as the latencies shown by infants and experimenters in the discrete trial-based task. The distribution shown in Figure 13 does not have such a clear peak around the 0.0 to -0.1 seconds bin. These results suggest that the discrete trial-based experiments constraint the variability in the latencies and reduces considerably the latency that it usually takes for an infant to engage in Coordinated Joint Attention during every day play with the parent.

Discussion

The ability to attend to objects that others are attending to is used on a daily basis in everyday contexts but researchers have measured infants' joint attention outside these naturalistic contexts, in trial-based experiments. A growing body of literature shows that the phenomenon elicited in the trial-based experiments differs from the phenomenon that occurs when parents engage with their infants in more naturalistic contexts of play. The contribution of the present work is the explicit and exhaustive investigation of the relation between individual differences measured in discrete trial-based experiments and individual differences in infant behavior shown in continuous parent-infant free play sessions. All the evidence strongly and consistently shows that the individual differences captured in the trial-based context for Joint Attention are not representative of infant behavior in the more naturalistic context for Coordinated Joint Attention. Furthermore, many of the sensory-motor behaviors measured in the context of lab trial-based tasks that are highly predictive of important outcomes were not stable throughout the entirety of the trial and were also not related in individuals across trial types in the experiment. In fact, the results show that the experimental task uses potentially non-valid first-look based measures and it drastically contrasts with the free flowing context in terms of the temporal dynamics of the phenomenon elicited and the degree to which the infant's whole body is engaged with the objects. This work challenges what we think we know about joint attention in infants, calling us to revise the causal links and theories build around JA, some of which have led to interventions that may be not as helpful as expected. The results also raise larger questions about the practice of science, and highlight current barriers limiting understanding of fundamental phenomena in developmental science.

How to do Science that Translates to Real-world Contexts

Experiments were created to isolate and control causes of behavior to locate mechanisms, but naturalistic contexts do not do this and thus there is no control of the additional variables that interact with the phenomenon of interest. Experiments isolate causes to find effect but they also face limits because the behavior they elicit in the specified research context is completely context-dependent and in some cases it may not actually be representative of the real world. This larger translation problem from the lab to the real world applies to many other disciplines including the medical sciences, biology, chemistry and even physics (Kessler & Glasgow, 2011; Rothwell, 2005). The evidence presented here illustrates the Disconnect phenomenon since the research context specified in the discrete trial-based Joint Attention experiment elicits behavior that has nothing to do with behavior occurring in real-world contexts. We challenge the overreliance of the field on clean and controlled experiments that hope to reduce complexity and in doing so they miss the phenomena and mechanisms that actually drive development.

In a Disconnect phenomenon, the variance captured by the clean and controlled experiment is unrelated to the variance that occurs in the real world, or in more naturalistic contexts. The different sources of variance for each task may underlie the lack of correlation. Figure 14 is a conceptual model that shows the factors that influence infant behavior across contexts. General factors that affect infant behavior include motor, cognitive, social, emotional and motivational aspects, and are usually the processes researchers are interested in measuring and studying. These infant processes can only be measured in a research context. For this reason, both infant processes and task characteristics that interact with each other as Figure 14 shows determine the infant behavior captured in a research context. One clear example that shows the interaction is the A-not-B error, which comes and goes depending on task properties that affect infants' memory such as the salience of the dwells in which objects hide (Smith & Thelen, 2003). The key difference between discrete trial-based contexts and continuous free play contexts is the number of task characteristics that are fixed to take an arbitrary value. Discrete trial-based experiments need to define more characteristics while the continuous free play session only defines a few and lets the interaction unfold as naturally as possible. Some of the task characteristics that need to be specified in the trial-based experiment but not in the free play session are crossed out from the list shown in Figure 14. As a result, the behavior elicited in the context of discrete trial-experiments may be more dependent on the context and its interaction with the infant's motor, cognitive, social, emotional and motivational processes that underlie behavior. The degree to which the behavior elicited by different contexts for research may represent more the actual infant processes and less the interaction with the context, depends on the ecological validity of the task characteristics and thus the behavior elicited. Researchers across the board, but especially those conducting experimental research, should actively think about the values that the task characteristics take and limit the contextual influences so that they do not study only a lab-based phenomenon that is far removed from the behavior elicited by more real-world research contexts.

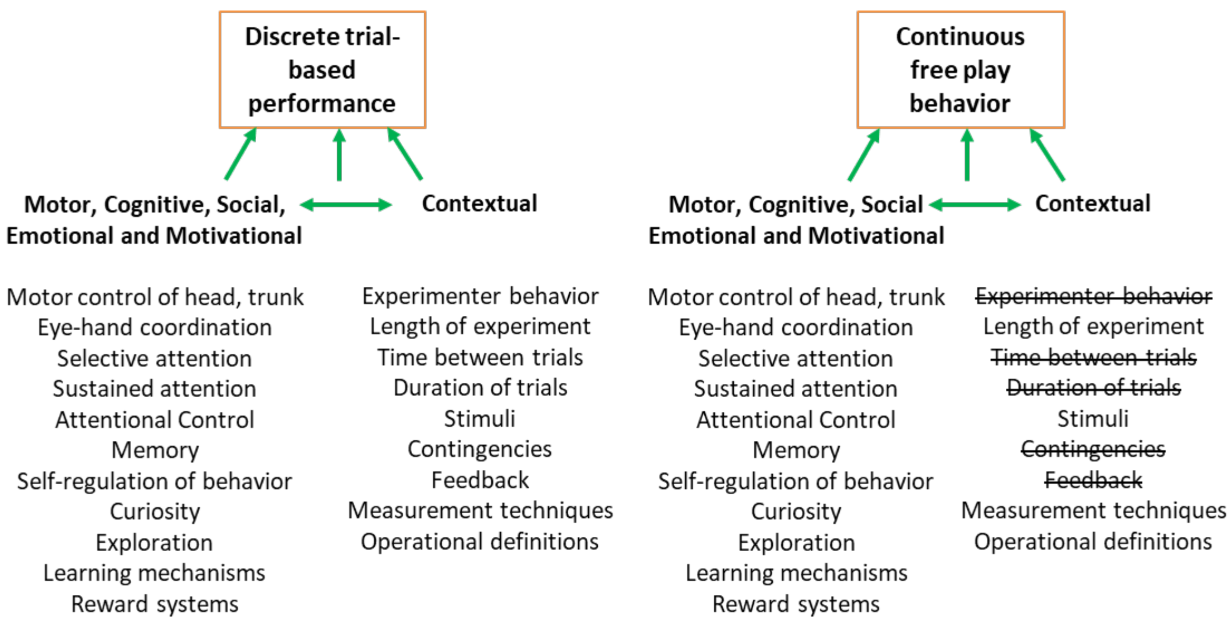


Figure 14. Determinants of discrete trial-based performance and behavior in continuous free play sessions.

Being aware of the contextual influences on infant behavior is not enough to solve the translation problem because ultimately, researchers are forced to make decisions and fix task characteristics in order to run their experiments. A researcher, for instance, will have to pick the duration of the trials and also decide to use a female versus a male experimenter. An experiment is the product of a series of decisions and it represents one out of the multiple possible combinations of the values that all the task characteristics could have taken. Figure 15 illustrates the process, where each node in the path represents a decision and thus each path created represents a different experiment that could be run. A researcher would have to run more than a dozen of experiments that take different task characteristics to tease apart the effects of the task characteristics from the infant processes of interest. In a study design that makes three different decisions for which there are only two options, the total number of possible designs is 14. Because researchers need to make decisions that result in many possible combinations but cannot test all possible combinations to investigate infant processes that are not dependent on the context, we propose researchers should consider moving away from experimental designs that may elicit behavior like JA, which is less representative of infant processes and more dependent on contextual factors.

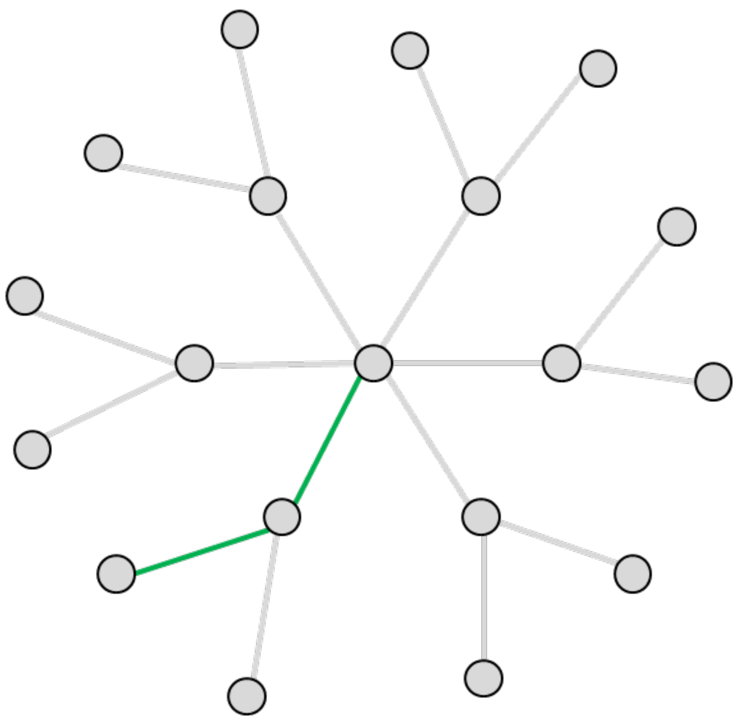


Figure 15. The pathway it took to define an experiment shown in green and in the midst of all possible experiments.

Contexts for research that are more naturalistic may elicit behavior that is less context dependent and more the result of the infant attention processes, but these contexts may also pose new challenges for researchers who are used to conduct research in the clean and controlled contexts. Every moment of the free flowing play is an experimenter's nightmare. Human behavior at every moment is influenced by infant processes that change dynamically as the interaction unfolds and thus the present moment depends also on the previous state of the infant in the previous time step. The real world never presents itself in the same way and an infant may never respond in the same way. Nonetheless, the complexity of the world matters because it is part of the training ground that supports development. Infants' everyday contexts train them in multiple pathways that result in social interaction and Coordinated Joint Attention because the world is full of moving parts and combinations of values. Accordingly, it is not enough to capture of the pathways that lead to phenomena in the context of an experiment. The beauty of the approach is actually seeing all pathways and mechanisms operating and interacting together in real time. Researchers are called to use event analyses and models to fully describe the behavior and related factors, model the values that variables take and find a strong signals that are recovered from the noise but are robust and valid.

Cascading Effects of the JA Disconnect

Research that aims to understand how children develop needs to relate to the real world because it is in the real world where development happens. Mechanisms discovered in the lab can actually harm science and society if they do not play out in the real world. Accordingly, the phenomenon elicited in the discrete trial-based context led researchers to theorize joint attention is about looking at the eye gaze of each other, interpret intentions and follow gaze. The big danger is that researchers may take the abilities that are demonstrated in these discrete trial-based tasks at the very surface level and assume these are the actual mechanisms or processes that support real development and behavior in the real world. This is exactly what occurred in our field with joint attention. The Joint Attention Disconnect is deeply problematic for our field because research carried out in the context of discrete trial-base experiments led to theories that claim JA is about understanding of minds and intentions which lead infants to learn language more easily. But the current work in addition to the growing body of literature suggests that actually the mechanisms of the lab phenomena are not likely to occur in everyday contexts (Yu & Smith, 2013; 2017a; 2017b) because actually the phenomena elicited in the discrete trial-based task has nothing to do with the behavior that occurs in more real-world contexts. Nonetheless, these wrong theories led to interventions that today train gaze-following skills in children diagnosed with autism, in order to improve their social interactions and language skills. The results from the work presented here show that gaze-following led to Joint Attention only as likely as hand-following did, and that the ability to follow gaze is not predictive of any behaviors of infants in free play, which would be the best indicator that the intervention is helping. The evidence may suggest gaze-following interventions target a skill that is not important but researchers and clinicians do not know this since the long-standing assumption is that joint attention occurs through gaze-following and is a fundamental mechanism that supports good social interactions and language learning. We propose that the measures obtained during discrete trials although not valid of the entire trial as the results show, should only be used as biomarkers, diagnostic tools but not as the mechanism itself that supports the outcomes they predict.

Why is discrete trial-based performance related to language learning then? One possibility is that both share sources of variance just as ice cream consumption is correlated with shark attacks but neither cause the other one. The temperature influences both phenomena, but each one is not a cause of the other one, they just share the same determining factor. Figure 16 shows that there are various overlapping factors that determine both discrete trial-based performance and future developmental outcomes. As it is clear, the quality of parent-infant interactions, joint attention and social cognition could support developmental outcomes but the other infant processes that actually affect discrete trial-based performance could support the outcomes. Shared sources of variance could be responsible for the apparent but perhaps *non-existent* causal association between discrete trial-based performance and developmental outcomes. Akhtar & Gernsbacher (2007) develop a strong argument that suggests Joint Attention is not causally related to word learning.

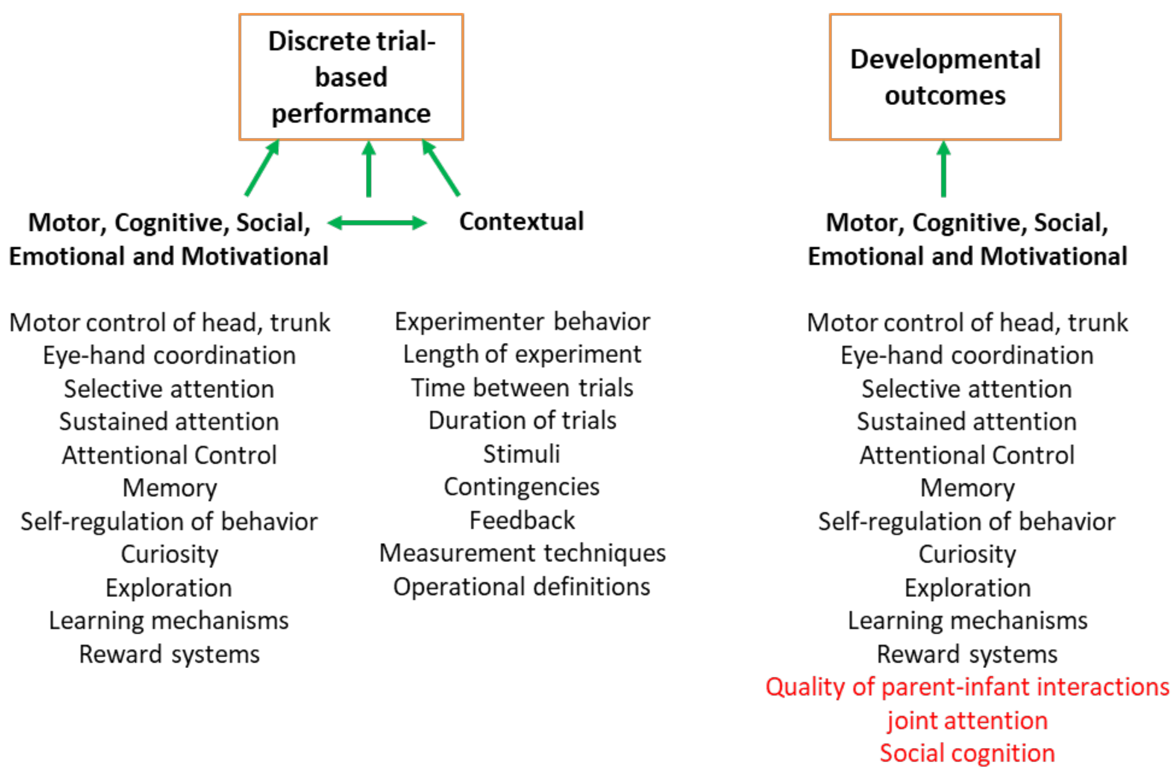


Figure 16. Determinants of discrete trial-based performance and developmental outcomes.

Another possibility is that there is a causal pathway that links behavior measured during trial-based experiments and language learning, but the pathway is at this point unknown because it is not clear what exactly the discrete trial-based experiment is measuring. Is the task capturing infants' abilities to disengage from an attractive stimulus? The evidence presented on the infant sensory-motor behaviors and the temporal dynamics may suggest that the task is almost like a disengagement task because infants are not actually playing with and touching objects but mostly attending to the objects and the face during the task. Furthermore, the shift in gaze from the face to the cued object by the experimenter occurs very rapidly at the level of 500-750 milliseconds, a reaction time that differs from the reaction time of play with the parent as shown in this paper, but that is similar to that obtained by researchers that study visual attention in response to labeling visual targets (Fernald, Perfors & Marchman, 2006) and disengagement from a central stimulus (Peltola, Leppänen, Vogel-Farley, Hietanen & Nelson, 2009). Would a task that flashes a light in the periphery of a screen that shows an engaging central stimulus, capture related variance?

Another possibility is that the discrete trial task measures infants' ability to recover from perturbations of the natural dynamics of play. The discrete trial-based task modified the natural characteristics of a social interaction in which joint attention takes place. The interaction was split into discrete trials, had an unfamiliar adult that did not respond to the infant's lead, and also made infants play a passive role, which created a movie watching feeling and prevented immediate reward and feedback. Future research is needed to investigate what exactly the standard gaze-following task measure since the present work suggests it does not measure infant joint attention, infant behaviors that support learning, or sensory-motor behaviors.

Current Challenges and Open Questions

The challenges that developmental scientists face is putting together all the mechanisms tested in controlled experiments and investigating whether they play out in the real world to bring behavior that drives developmental change. Some phenomena may actually fail to translate such as mutual exclusivity and the A-not-B error but those that will translate, will generate valid and useful knowledge about how children develop. The endeavor is promising since it may stop researchers from studying lab-based phenomena that do not explain valid behavior. Moreover, it may result in the specification of phenomena as they occur and interact in the real world. The new way of doing basic science should refrain from aiming to establish causal with each study and from trying to make simple what is complex, which could lead to never really comprehending development. Researchers need to embrace the complexity of the world and the complex pathways that actually matter for developmental change.

Open questions for the investigation of Coordinated JA include the moment-to-moment consequences of the phenomenon for key infant processes including sustained attention and developmental outcomes such as language learning, self-regulation and academic achievement. The two remaining chapters in this dissertation address this open question by examining the effect of Coordinated JA on infants' in-the-moment (Chapter 3) and future (Chapter 4) control of sustained visual attention. Together these papers leverage on the examination of infant and parent behavior moment-to-moment in the context of naturalistic play session and build a path from social partners to future developmental outcomes through sustained attention.

Additionally, research is needed to understand how Coordinated JA occurs in even more naturalistic contexts of the home in order to examine the everyday behaviors that infants use to engage in joint attention with a social partner that has competing attentional demands and transitions between daily routines that include play as well as feeding and grooming (Tamis-LeMonda et al., 2018). There is research showing that although free play sessions that take place in the lab are representative of behavior in the home and individual differences are correlated

(Tamis-LeMonda et al., 2016), the lab context still changes the behavior of parents and their infants. In particular, experimenters may choose undefined instructions that result in unnatural behavior and infants may display less disruptive behaviors while parents may display more sensitive behavior (Gardner, 2000). In fact, the test-retest reliability of the measurements obtained in the free play paradigm used throughout this dissertation has not been calculated. As a first step, we examined test-retest reliability by splitting every play session in half and compared measures of infant and parent behaviors (see Supplemental material) at the subject level. The results showed that infant and parent behavior was relatively constant between the first and second halves, but more work is needed to actually bring dyads to the lab on separate days and repeat the same play session to more cleanly assess the test-retest reliability of the measures obtained.

Finally, research is needed to understand infant and parent dynamics at the physiological level obtaining heart rate and brain measurements while they engage in everyday interactions with moments of Coordinated JA. According to the results presented here, Coordinated Joint Attention seems to occur behaviorally through gaze-following, hand-following or a still unknown pathway, but are there physiological correlates that separate these pathways to engage in joint attention?

In conclusion, experiments are not bulletproof because they can locate mechanisms and study behavior that has nothing to do with the mechanisms and behaviors that actually play out in the real world. The research context is the product of a hierarchy of decisions, and each decision is nested within another. For any experiment, the context that was specified for the study could set task characteristics that may perturb the system and processes that under different research contexts would generate different behavior; but researchers may never intend to do this.

Perhaps those unnatural task characteristics were chosen because they were practical and actually made the experiment possible. Tasks could also be defined in ways that make the experiment standard across subjects. This paper presented evidence of another instance of the Disconnect phenomenon and proposed that if we want to understand how joint attention occurs in the real world between social partners and infants as well as how it helps for development, discrete trial-based experiments will not tell us anything. Development occurs as the result of the behavior of a tightly coupled system at various timescales that moves together and is highly influenced by the context. Research needs to move away from clean petri dish designs; we need a new basic science one that focuses on examining complexity and study valid behavior.

Supplemental Material

Silent and Non-silent Trials

Half of the Gaze trials had a naming utterance accompanying the gaze cue and half did not, but JA during the cue did not differ between the two types of Gaze trials in terms of the proportion of time the infant and adult spent looking at the target object ($t(26)=1.01$, $p=0.32$). Gaze trials with and without a naming did not differ as well in the duration of the Joint Attention episodes ($t(26)=1.61$, $p=0.12$) and in the frequency per minute of Joint Attention episodes generated during the cue ($t(26)=-0.01$, $p=0.99$).

In the same way, half of the Gaze-Hold trials that had a naming utterance accompanying the cue did not differ from the Gaze-Hold trials that did not have a label in terms of the proportion of time spent in Joint Attention ($t(26)=0.43$, $p=0.67$), the duration of the Joint

Attention episodes (t(26)=1.97, p=0.06) and the frequency of Joint Attention episodes (t(26)=-1.46, p=0.16).

Predicting Coordinated JA from Multiple Measures from the Discrete Trial-based Experiment

Table 12

Model 1: Predicting Coordinated JA from 7 measures obtained in Gaze trials

Coefficients	B	SE	T	p
Intercept	0.39	0.19	2.08	0.05
Mean prop. of Gaze trial time looking at face	-0.32	0.16	-1.95	0.07
Prop. of Gaze trials with at least one JA	-0.29	0.20	-1.46	0.16
Prop. of Gaze trials with at least one face look and JA	0.39	0.21	1.83	0.08
Mean duration of first JA in Gaze trial	-0.02	0.03	-0.59	0.57
Mean latency to first JA in Gaze trial	-0.01	0.04	-0.27	0.79
Prop. of Gaze trials with look to face then target	-0.13	0.10	-1.22	0.24
Looking score for Gaze trials	0.01	0.01	0.60	0.55
R^2 = 0.24; Adjusted R^2 = -0.04; F(>19) = 0.87; p = 0.55				

Table 13

Model 2: Predicting Coordinated JA from 5 measures obtained in Gaze-Hold trials

Coefficients	B	SE	T	p
Intercept	0.14	0.23	0.61	0.55
Mean prop. of Gaze-Hold trial time looking at target	-0.03	0.36	-0.08	0.93
Mean duration of first JA in Gaze-Hold trial	0.00	0.03	0.09	0.93
Mean latency to first JA in Gaze-Hold trial	0.18	0.08	2.19	0.04*
Prop. of Gaze-Hold trials with look to face then target	0.06	0.08	0.68	0.51
Looking score for Gaze-Hold trials	0.01	0.01	0.87	0.39
R^2 = 0.22; Adjusted R^2 = 0.04; F(>19) = 1.21; p = 0.34; *=p<0.05				

Table 14

Model 3: Predicting parent-led Coordinated JA from 7 measures obtained in Gaze trials

Coefficients	B	SE	T	p
Intercept	0.09	0.09	1.05	0.31
Mean prop. of Gaze trial time looking at face	-0.04	0.08	-0.49	0.63

Prop. of Gaze trials with at least one JA	-0.05	0.09	-0.57	0.58
Prop. of Gaze trials with at least one face look and JA	0.10	0.10	1.01	0.33
Mean duration of first JA in Gaze trial	0.00	0.01	-0.05	0.96
Mean latency to first JA in Gaze trial	0.00	0.02	-0.05	0.96
Prop. of Gaze trials with look to face then target	-0.06	0.05	-1.17	0.26
Looking score for Gaze trials	0.00	0.00	1.00	0.33
$R^2 = 0.15$; Adjusted $R^2 = -0.16$; $F(>19) = 0.47$; $p = 0.84$				

Table 15

Model 4: Predicting parent-led Coordinated JA from 5measures obtained in Gaze-Hold trials

Coefficients	B	SE	T	p
Intercept	0.14	0.11	1.26	0.22
Mean prop. of Gaze-Hold trial time looking at target	-0.18	0.17	-1.06	0.30
Mean duration of first JA in Gaze-Hold trial	0.01	0.01	1.21	0.24
Mean latency to first JA in Gaze-Hold trial	0.01	0.04	0.17	0.87
Prop. of Gaze-Hold trials with look to face then target	0.03	0.04	0.69	0.50
Looking score for Gaze-Hold trials	0.00	0.00	1.06	0.30
$R^2 = 0.13$; Adjusted $R^2 = -0.08$; $F(>19) = 0.63$; $p = 0.68$				

Table 16

Model 5: Predicting parent-led Coordinated JA through Gaze following from 7 measures obtained in Gaze trials

Coefficients	B	SE	T	p
Intercept	0.29	0.18	1.60	0.13
Mean prop. of Gaze trial time looking at face	-0.05	0.16	-0.29	0.77
Prop. of Gaze trials with at least one JA	-0.03	0.19	-0.18	0.86
Prop. of Gaze trials with at least one face look and JA	-0.10	0.21	-0.47	0.65
Mean duration of first JA in Gaze trial	-0.01	0.03	-0.44	0.66
Mean latency to first JA in Gaze trial	-0.05	0.04	-1.07	0.30
Prop. of Gaze trials with look to face then target	0.03	0.10	0.29	0.78
Looking score for Gaze trials	0.00	0.01	0.38	0.71
$R^2 = 0.19$; Adjusted $R^2 = -0.11$; $F(>19) = 0.64$; $p = 0.72$				

Table 17

Model 6: Predicting parent-led Coordinated JA through Hand following from 5 measures obtained in Gaze-Hold trials

Coefficients	B	SE	T	p
Intercept	-0.04	0.24	-0.19	0.85
Mean prop. of Gaze-Hold trial time looking at target	0.20	0.36	0.56	0.58
Mean duration of first JA in Gaze-Hold trial	-0.01	0.04	-0.34	0.74
Mean latency to first JA in Gaze-Hold trial	0.12	0.08	1.45	0.16
Prop. of Gaze-Hold trials with look to face then target	0.04	0.08	0.43	0.67
Looking score for Gaze-Hold trials	0.01	0.01	1.06	0.30
R^2 = 0.11; Adjusted R^2 = -0.10; F(>19) = 0.51; p = 0.77				

Reliability of the Free Play Session Measures

The following behaviors were chosen to examine the test-retest reliability of the measurements obtained from infant and parent behavior during the free play session: infant looking at the four ROIs, parent looking at the four ROIs, Coordinated JA, infant-led Coordinated JA, parent-led Coordinated JA, parent talk, infant manually holding objects and parent manually holding objects. Each subjects' proportion of playtime spent in displaying each of these behaviors was calculated for the first half of the play session (trials 1 and 2 for 25 subjects, and only trial 1 for 2 subjects) and for the second half of the play session. Figure 17 shows the mean of these subject-level-calculated proportions of playtime spent in each of these behaviors for the first and second half of the play session. Table 18 shows the means, standard deviations and standard errors. The figure and table show that parents and infants spent the same proportion of playtime in most of these behaviors. When they differed, behavior occurred in greater proportions during the first half of the experiment but the difference did not seem to be any larger than the variance shown across subjects within the first and second half.

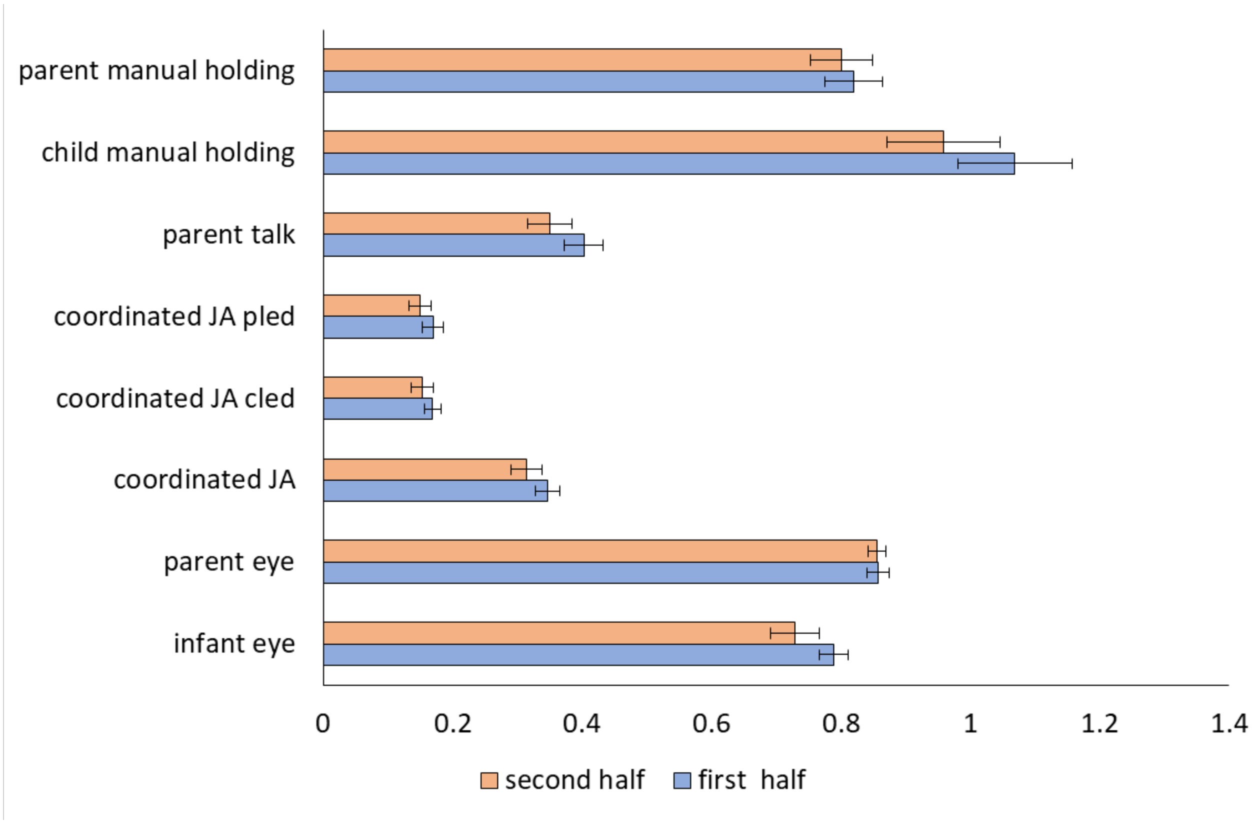


Figure 17. Proportion of playtime spent in the different parent and infant behaviors compared during the first and second half of the play session.

Table 18.

Measures of Parent and Infant Behavior Calculated for the First and Second half of the Play Session

	first half			second half		
	mean	SD	SE	mean	SD	SE
infant eye	0.79	0.12	0.02	0.73	0.20	0.04
parent eye	0.86	0.09	0.02	0.86	0.07	0.01
coordinated JA	0.35	0.10	0.02	0.31	0.13	0.02
coordinated JA c-led	0.17	0.07	0.01	0.15	0.09	0.02
coordinated JA p-led	0.17	0.08	0.02	0.15	0.09	0.02
parent talk	0.40	0.16	0.03	0.35	0.18	0.03
child manual holding	1.07	0.46	0.09	0.96	0.45	0.09
parent manual holding	0.82	0.23	0.04	0.80	0.25	0.05

Chapter 3

Multimodal Parent Behaviors within Joint Attention Support Sustained Attention in Infants

Perceivers sometimes glance at objects but other times they visually examine a single object with a sustained look. These longer looks are strongly related to visual learning about the attended target, in infants (Lansik & Richards, 1997; Richards & Casey, 1992; Ruff, 1986), young children (Ruff & Lawson, 1990) and adults (Steinmayr, Ziegler & Träuble, 2010; Wei, Wang & Klausner, 2012). Sustained visual attention develops incrementally from late infancy through early childhood; for example, the average time that a one-year-old infant attends to a single toy during active play is about 3 seconds whereas the average time for a 3-year-old child approaches 9 seconds (Ruff & Lawson, 1990). Early individual differences in sustained attention to objects predict later individual differences in inhibitory control and self-regulation (Kochanska, Murray, & Harlan, 2000; Reck & Hund, 2011; Ruff, 1986), language (Welsh, Nix, Blair, Bierman & Nelson, 2010), and school achievement (Duncan, Dowsett, Claessens, Magnuson, Huston, Klebanov, Pagani, Feinstein, Engel, Brooks-Gunn, Sexton, Duckworth & Japel, 2007; McClelland, Acock & Morrison, 2006; McClelland & Cameron, 2012). The factors that drive the development of sustained attention and their role in the observed individual differences in sustained attention have not been identified. The ability to sustain attention is sometimes treated as an intrinsic and stable child attribute related to temperament (Colombo, 2001; Posner & Rothbart, 2000). However, developmental theorists have also suggested that parents and social context play a role (e.g., Kopp, 1982; Miller, Ables, King & West, 2009; Parrinello & Ruff, 1988; Sigel, 2002; Spruijt, Dekker, Ziermans, & Swaab, 2018; Vygotsky, 1978). The potential role of parent behaviors in infant sustained attention is the focus of this research.

Yu and Smith (2016) recently showed a direct social effect on the duration of infant attention to a single object during toy play with a parent. Their study used dual head-mounted eye trackers to measure both parent and infant (12 month olds) gaze during play. They defined *infant sustained attention* as an unbroken look to an object that was longer than 3 sec. They defined Coordinated Joint Attention objectively, as moments in which the two participants' gaze was directed to the same object, without considering how that joint gaze was achieved and without regard to what one might infer about the knowledge states of the participants, a definition of joint attention distinct from that used in the past (Baron-Cohen & Cross, 1992; Mundy, Sullivan & Mastergeorge, 2009; Tomasello, 1995). The main result was that toddler *sustained* attention occurred most frequently when the period of infant attention to an object overlapped with a Coordinated *Joint* Attention episode. Yu and Smith suggested that parent attention to an infant-attended object may extend the infant's own interest, causing the infant to visually attend to the object longer than the infant would otherwise. Yu and Smith further suggested that this in-the-moment extension of the duration of infant attention by parent shared interest –when repeated day-in and day-out – may tune and train the internal mechanisms that support the development of the self-regulation of attention. They offered the following analogy from how children learn to ride two-wheel bikes to explain the idea: parents often hold onto and balance a two-wheel bike for the young rider at the beginning, so that the learner can get the feel of balancing a bike. After repeated episodes of such parent support, the child becomes able to balance a two-wheeler without help. In the same way, Yu and Smith proposed, parent joint attention to an object may help infants stay attending to that object and through these socially-guided moments of sustained attention events, infants may develop the means to sustain focused attention on their own.

There are many untested predictions that follow from this hypothesis. Here we focus on one open question relevant to the future testing of those hypotheses: What is happening inside Coordinated Joint Attention episodes that supports the infant's longer looks to an object? This question is critical because the only parent behavior measured by Yu and Smith was parent gaze in coordination with infant gaze. In principle, the child's extended attention to the object –given joint parent attention to that object– could result from the infant perceiving the direction of parent gaze and inferring parent shared interest in the object from gaze direction alone (e.g., Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005). However, Yu and Smith also reported that during play with toys, infants almost never looked to their parents' face and thus couldn't use parent eye-gaze to infer that the parent was also looking at the same object. The finding that toddlers do not often look to the parent's face during joint object play, has been previously reported by multiple laboratories (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013). Accordingly, the present study was designed to address two empirical questions: (1) what additional parent behaviors are part of Coordinated Joint parent-infant Attention to an object? (2) Are these additional parent behaviors associated with longer visual attention to the object attended by the infant?

When parents are playing with their infants in free-flow interactions, they may do more than just look at objects and at their infants. Parents generate multimodal behaviors and their own attention and interest in the object is potentially signaled through multiple modalities including handling objects, talking about the objects as well as looking at those objects

(Bakeman & Adamson, 1984; Tomasello & Farrar, 1986; Yu & Smith, 2012; Yu & Smith, 2013). Several researchers have specifically observed that parent talk increases infant visual attention to objects (Baldwin and Markman, 1989; Belsky, Goode and Most, 1980; see also Parrinello & Ruff, 1988). Other evidence suggests that parent hand actions also play a role in organizing infant visual attention to objects (Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Yu & Smith, 2013, 2017). Accordingly, we tested the hypothesis that multimodal parent behaviors, such as parent talk and parent handling of objects, are often components of a Coordinated Joint Attention episode that are associated with longer lasting visual attention to the object by the infant. Because parent looks, utterances and manual activities during toy play are real-time behaviors happening at the time scales of seconds and fractions of seconds, our analyses also focus on the finer temporal details of parent behaviors and their real-time effects on infant gaze.

Method

Participants

The participants were 40 infants with a mean age of 13.97 months (range 12.1 to 16.1 months, 19 females) and their parents. There was a failure of auditory recording on one infant whose data were included in analyses of infant gaze distributions and the relation of infant gaze to parent gaze and touch, but were excluded in analyses involving parent talk. Eighteen additional infants were recruited but did not contribute data because of refusal to wear the head-mounted device or other technical failure. Families were recruited from a population of working and middle class families. Participants were given a small toy as compensation for their participation in the study. This research project was approved by the Research Subjects Review

Board at Indiana University (protocol number 0808000094) and titled “Multimodal word learning”.

Stimuli

A pool of 15 novel objects (on average, about 9.50 x 6.5 x 5.0 cm) was created in the lab with unique shapes and objects were combined in sets of 6 in three different ways (each set defined a unique experiment because the selected participants for the analyses were pooled from these three experiments). Through piloting, objects were designed to be visually and manipulatively engaging. All children were given objects in sets of three in which one was painted blue, one red and one green. Each child played with two unique sets of objects and the criterion for selecting a set of toys for an infant was that the toys were novel for that infant. Figure 1 shows a participating dyad as well as one of the sets of objects used for the study.

Experimental Room

Parent and infant sat across from each other at a white table 61cm × 91cm × 64cm (see in Figure 1). The infant sat on a chair and the parent sat on the floor such that the tabletop was approximately 46 cm from the center of the table to the child’s eye and to the parent’s eye. Both participants wore head- mounted eye trackers (Positive Science LLC, http://www.positivescience.com; also see Franchak & Adolph, 2010; Franchak et al., 2011). Both parent and infant eye-tracking systems included an infrared camera –mounted on the head and pointed to the right eye of the participant that recorded eye images– and a scene camera that captured the events from the participant’s perspective. The scene camera’s visual field was 108 degrees. Each eye tracking system recorded both the egocentric-view video and eye-in-head position (x and y) in the captured scene at a sampling rate of 30 Hz. In order to eliminate distractors in the environment and encourage infants to focus on object play, everything in the room –other than the objects and the hands and faces of the participants- was white. Three additional cameras recorded the interaction from third-person views.

Figure 1. Parent and infant in the tabletop with novel objects and dual head-mounted eye-tracking. The authors received signed consent for the parent and infant's likenesses to be published in this dissertation.

Procedure

Prior to entering the testing room, an experimenter desensitized the infant to touches to the head and hair by lightly touching the hair several times when the attention and interest of the infant was directed to a toy. Upon entering the experimental room, a second experimenter and the parent engaged the infant with a toy with buttons to push that made animals pop up as the first experimenter placed the head gear on the infant. This was done in one movement and care was taken to ensure that the infant remained engaged with the toy and that the infant's hands did not go to the headgear. The first experimenter then adjusted the scene camera to ensure that the button being pushed by the infant was in the center of the scene camera. The experimenter then directed the child's attention toward an attractive toy on the table ensuring the child's eyes were following the toy. This procedure was repeated in 15 different locations on the tabletop to ensure a sufficient number of calibration points for the infant's eye-tracking. After placing the parent's head gear the experimenter asked the parent to look at one of the objects on the table in various locations. This procedure was repeated 15 times in order to obtain at a sufficient number of calibration points for the parent's eye-tracker. Parents were instructed to play with their child with 3 toys at a time as they would normally do at home and asked if they named the object to use the objects' labels provided. Parent-infant dyads played in four 1.5-minutes-long trials, using two different sets of 3 toys in an alternating fashion across the 4 trials. The duration of the trials was chosen so that infants remained engaged in play with limited off-task behavior during the entire experiment. This duration is also consistent with previous research (Yu & Smith, 2013; Yu & Smith, 2016; Yu & Smith, 2017).

Coding and Definitions

The quality of the eye tracking videos (with eye images superimposed) for each infant and parent was checked (using centered hand actions on an object as described above) to ensure the quality of calibration throughout the session, at the end as well as at the beginning of the session. The eye-tracker collected data at a rate of 30 frames per second for approximately 360 seconds (four trials with 1.5 minutes per trial) of interaction, yielding potentially 10,800 data points per measure for each participant. Of this total possible, the number of analyzed frames – frames in which infant gaze was directed to one of the regions of interest—were 7,816 (SD = 1,893) and for parents, it was 8,697 (SD= 1,970). The "missing" frames include eye-blinks and periods when the infant was off-task (e.g. looking around the room rather than at the objects or parent).

Looks. The main data for analyses were eye gaze data directed to four regions of interest (ROIs): the three objects in play at any time and the partner's face. ROI coding was done by highly trained human coders who continuously code these variables for multiple experiments without knowledge of or regard to the hypotheses under test. The ROI coding for this experiment was done as part of that workflow. Each ROI was strictly defined in terms of the in-view pixels belonging to the object. The coders annotated gaze direction –frame by frame – judging whether the cross hairs fell on the pixels of the ROI. Thus, an *unbroken look* to an object might have multiple fixations on the object as long as all gaze fell within the ROI. Reliability was computed between the coding of two independent coders on eleven dyads that were randomly selected. Coders coded 25% of each dyad's frames making judgments on 2,790 frames per dyad on average. The inter-coder reliability of eye-gaze coding performed by these highly trained coders ranged from 82% to 95% assessed by Cohen's kappa of 0.75 (ranging from 0.57 to 0.91). This level of reliability is consistent with the reliability reported by previous published research (see Yu & Smith, 2016; Yu & Smith, 2017).

The main dependent variable is the duration of infant unbroken looks within an ROI. Infants may generate multiple fixations on the same object which is counted here as one unbroken look. For the reported analyses, the duration of an infant's continuous gaze within an

ROI needs to be longer than *500 milliseconds (msec) for that infant's continuous gaze* to be counted as a look. We did this because our interest is in sustained attention and because the dynamics of infant looking behavior are much slower than adults such that meaningful looks are at least this long (Yu & Smith, 2013; Yu & Smith, 2017). This approach allowed us to measure the parent behaviors that overlapped with unbroken and meaningful infant looks that were longer than this minimal duration. In order to ensure that this imposed floor on infant unbroken looks did not determine the results, we repeated the analyses using all infant unbroken looks, including those as brief as one frame (33 msec). The patterns of results and conclusions remained the same.

The main context of interest to examine infant looks is whether an infant look did or did not include joint attention with the parent, or moments in which the infant and parent looked at the same object. We defined a Coordinated Joint Attention objectively in terms of parent looks to an object that temporally overlapped with unbroken infant look to that same object. We counted all cases in which a parent look to the object (regardless of the duration and thus in principle as brief as a single frame or 33.3msec) overlapped with an infant look to that same object (defined as lasting at least 500msec). In this way, we divided all infant looks into two categories – those that overlapped with a parent look to the same object or those that did not. We took this approach so as to cleanly capture all looks by the infant of which the parent might be aware of the child's interest and thus behave in some way that encouraged that interest. Adults (unlike infants, Yu & Smith, 2013; Yu & Smith, 2017) rapidly shift gaze and can pick up and process useable information in very brief glances (Carpenter, 1988; Land & Hayhoe, 2001; Land, Mennie & Rusted, 1999).

Parent hand contact. Parent manual contact with an object was coded frame-by-frame from images captured by the overhead camera and the other two third-person cameras. Although parent touch was coded in all the frames, the only relevant coded frames for parent touch used for the analyses were frames that occurred when the infant looked at the objects. Parent touch was counted only when the parent touched the object attended by the infant. We used a custom coding program that allowed coders to access three views simultaneously to determine which object was manually handled frame by frame. In practice, coders most often relied on the overhead camera, but in cases of uncertainty could consult the other two views. Coders made frame-by-frame yes/no decisions that a parent hand was in contact with an object. A second coder also independently coded a randomly selected 25% of the frames of five parents and obtained inter-coder reliability assessed by Cohen's kappa of 0.90 (ranged from 0.76 to 0.96).

Parent talk. Parent speech was objectively coded at the utterance level, starting a new utterance after 400 milliseconds of silence (Suanda, Smith & Yu, 2016; Pereira, Smith & Yu, 2014; Yu & Smith, 2012). We included as speech, all sounds (words and word-like sounds) that included a vowel. This criteria excludes sounds such as coughs, raspberries or sighs and does not consider the content of the talk, treating naming ("it looks like a helicopter"), pointing out attributes ("that can spin") and general comments ("cool" or "that's fun") as all the same. Our assumption was that if parent talk had an effect on the duration of infant attention, it would be discernible from this coarse coding and thus our approach would be the right first step prior to a closer examination of effects of kinds of content, as well as prosody, which may be influential factors.

Statistical Analyses

The main empirical question concerns different kinds of Coordinated Joint Attention experiences based on their multimodal components (parent talk and touch) and their effects on infants' sustained attention to the jointly attended object. To this end, the results consist of three parts, with the first two being preliminary to the main question. First, we measure the duration of visual attention in infants. Second, we compare infant looks that did or did not include joint attention by the parent. By doing so, we replicate Yu & Smith (2016), testing the contribution of joint attention –that is parent look to the infant attended object– to the duration of the infant's unbroken visual attention to that object. Third, we turn to the main question, the multimodal nature of these joint attention episodes —defining four different categories — based on the combinations of parent talk and touching of the jointly attended object and their association with different durations of infant look to the object.

The main dependent variables of interest for all analyses are the durations of infant looks under different conditions. The distribution of infant look durations is extremely skewed (see Figure 2) as is true of many human behaviors (Clauset, Shalizi & Newman, 2009; Clerkin, Hart, Rehg, Yu & Smith, 2017; Kello, Brown, Ferrer-i-Cancho, Holden, Linkenkaer-Hansen, Rhodes & Van Orden, 2010; Piantadosi, 2014). Accordingly, and as is appropriate for these right skewed distributions, we categorized infant looks based on their durations into bins (e.g., brief, long and very long as defined below) and counted the number of looks for each participant in each bin (normalized as a proportion of all looks by that infant that were in each bin since different children had different numbers of looks) using both parametric and nonparametric statistics. The main analysis, however, consists of a linear mixed effects model with fixed and random effects (R Development Core Team, 2016) on the logs of the look durations.

Results

I. The Distribution of Infant Attention

Table 1 provides a summary of all infant unbroken looks to an object or parent face without regard to parent looks or other behaviors. Infants looked at the objects much more than they looked at the parent's face, a result that has been reported by many other investigators in a variety of social contexts for infants this age (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013). Figure 2 shows the histogram of the durations of infant unbroken looks to the objects, the main dependent measure in subsequent analyses. Infant look durations varied from 500 msec (the imposed floor) to nearly 31 sec. Most looks to objects were very brief but the tail of the distribution is quite long. By hypothesis, these very long but relatively infrequent looks that comprise the tail of the distribution are most relevant to assessing the role of parent behavior in sustaining infant attention. Accordingly, for the categorical analyses of look durations as a function of parent behavior, we used two main categories of durations: Brief looks, less than 3 sec (the threshold for sustained attention used in previous studies, Yu & Smith, 2016; Ruff & Lawson, 1990), accounted for about 75% of all infants looks and Long looks, 3 sec and longer, and typically considered sustained attention, accounted for about 25% of all infant looks. Within the Long looks –and *included in all Long Look analyses* -- we provide additional information about what we call Very Long looks. These are looks that are 10 sec and longer. They are not common, accounting for just 2% of all infant looks (Figure 2). Whereas all infants had at least some Long looks, not all of them had Very Long (10 sec or greater) looks. Nonetheless, we include results of Very Long looks because 65% of all infants had at least one Very Long Look, and because, as we report subsequently, these Very Long infant looks were associated exclusively with joint attention or moments in which the parent also looked at the infant-attended object.

Table 1

Descriptive Statistics of Infant Looking Behavior to Objects and Parent Face

Infant looks	Mean prop of infant looks	SD prop of infant looks	Median prop of infant looks	Range in prop of infant looks	Mean duration (sec)	SD duration (sec)	Median duration (sec)	Range in duration (sec)
At objects	0.78	0.07	0.75	0.64 – 0.97	2.40	0.51	2.26	1.60 – 3.82
Parent face	0.22	0.07	0.25	0.03 – 0.36	1.79	0.51	1.73	0.90 – 3.61

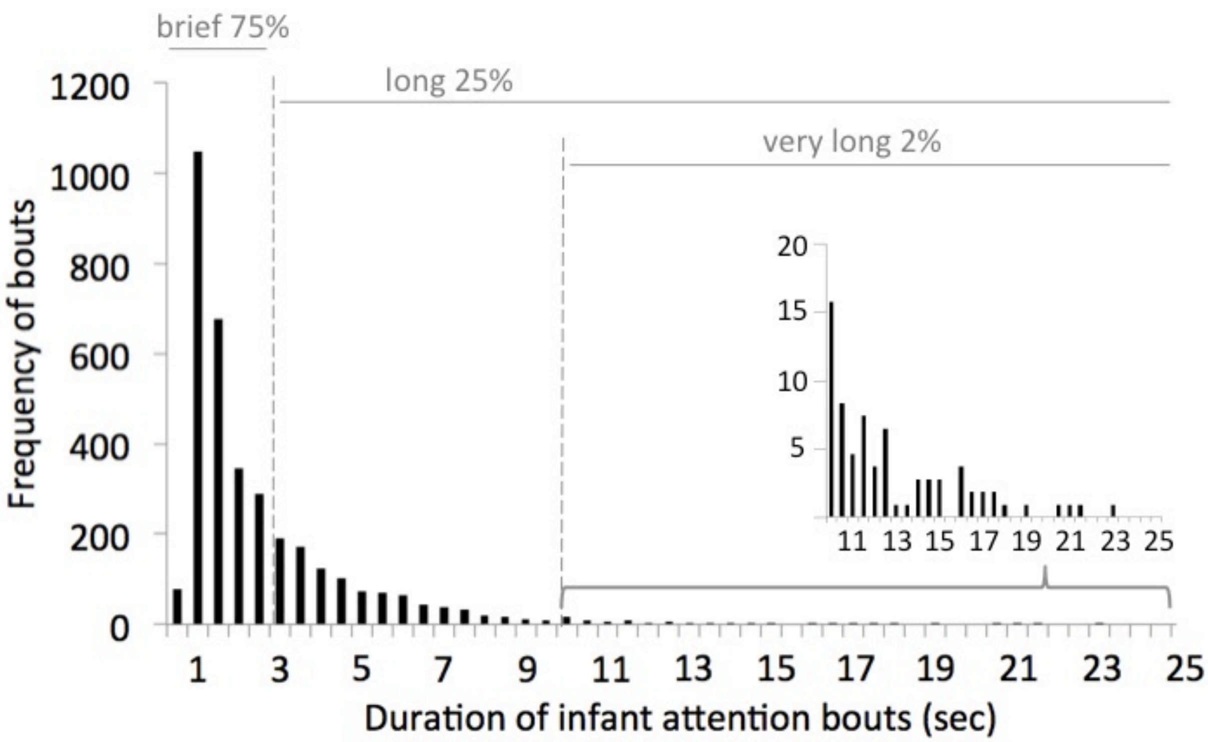


Figure 2. Frequency distribution of the duration of infant attention bouts to objects measured in seconds. Percentage of brief, long and very long looks present in the distribution are shown.

II. Coordinated Joint Visual Attention

Each infant look was classified in one of two mutually exclusive categories–including or not including Coordinated Joint Attention with the parent. If during any period of the infant's continuous gaze to the object, the parent also directed gaze to same object (no matter how briefly), *the entire infant look* was categorized as including joint attention, as illustrated in Figure 3. The durations of parent overlapping looks could be short or long, and the parent look to the object could follow the infant's look (infant-led) or could precede it (parent-led). By these definitions, parent and infants jointly attended to objects during play 49% (SE=2%) of the play session whereas infants looked at objects without an overlapping parent look 12% (SE=1%) of the time. The remaining times consists of looks shorter than 500 msec, looks elsewhere in the room, or to the partner's face. Overall, the results show, as has been reported before (Yu & Smith, 2013) that parents and infants consistently coordinate their gaze to the same object during free-flowing play.

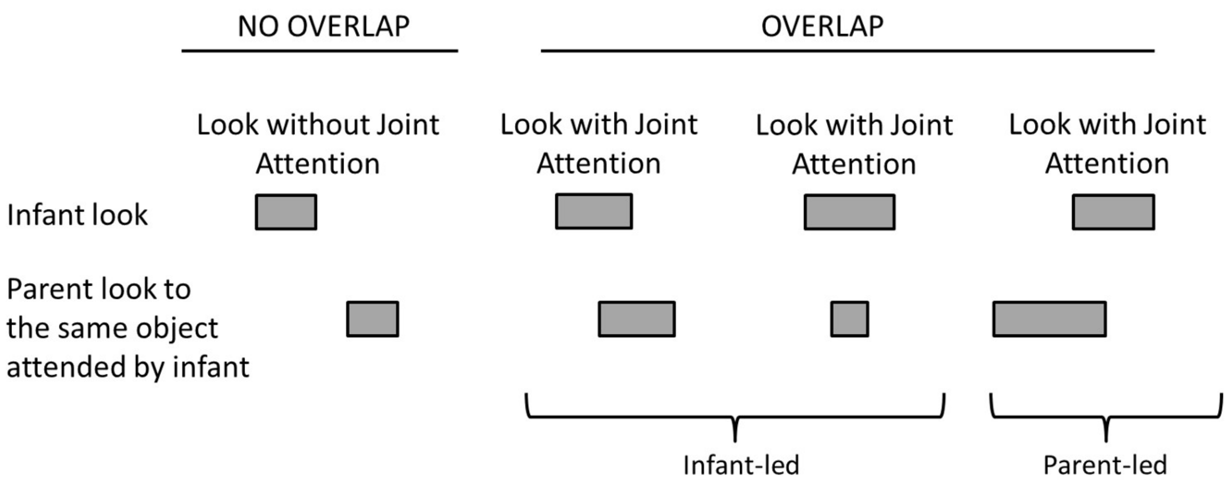


Figure 3. Definitions for infant's look without Coordinated Joint Attention and with Coordinated Joint Attention (both infant-led and parent-led cases) based on overlap with a parent's look. Joint Attention was defined objectively as the temporal overlap between an infant's look at an object and the parent's look at the same infant-attended object.

Figure 4A shows the histogram of the durations of the parents' overlapping looks to the same object to which the infant attended. Most parent looks were very brief and overall much shorter than infant looks (parent looks were on average 1.28-seconds-long (SD=0.53, Median=1.16), infant looks were on average 2.26-seconds-long (SD=0.51, Median=2.26)). Figure 5A shows the frequency of parent-led and child-led looks to the same object and the relative timing of the onset of parent looks to the infant-attended object. We used very small bin sizes incrementing at a tenth of a second in order to show the tight temporal coordination of parent-and infant gaze to the same object (see also Yu and Smith, 2013). Parents follow the infant's gaze to the object more often than they lead, but they follow rapidly; the mode gap between infant onset and parent onset of gaze to the same object is at the bin between 0 and .10 second. Figure 5A shows that 76% of the overlapping parent looks occurred within 1 second before or 1 second after the onset of the infant's look. This tight coordination in time rules out one uninteresting account of why infant looks might be longer when parents also look at the same object: longer looks by the infant could have provided more time for an overlapping glance by the parent and thus greater likelihood the infant look is counted as including joint attention. But this did not happen as parents looked to the object close in time to the start of the infant's own look.

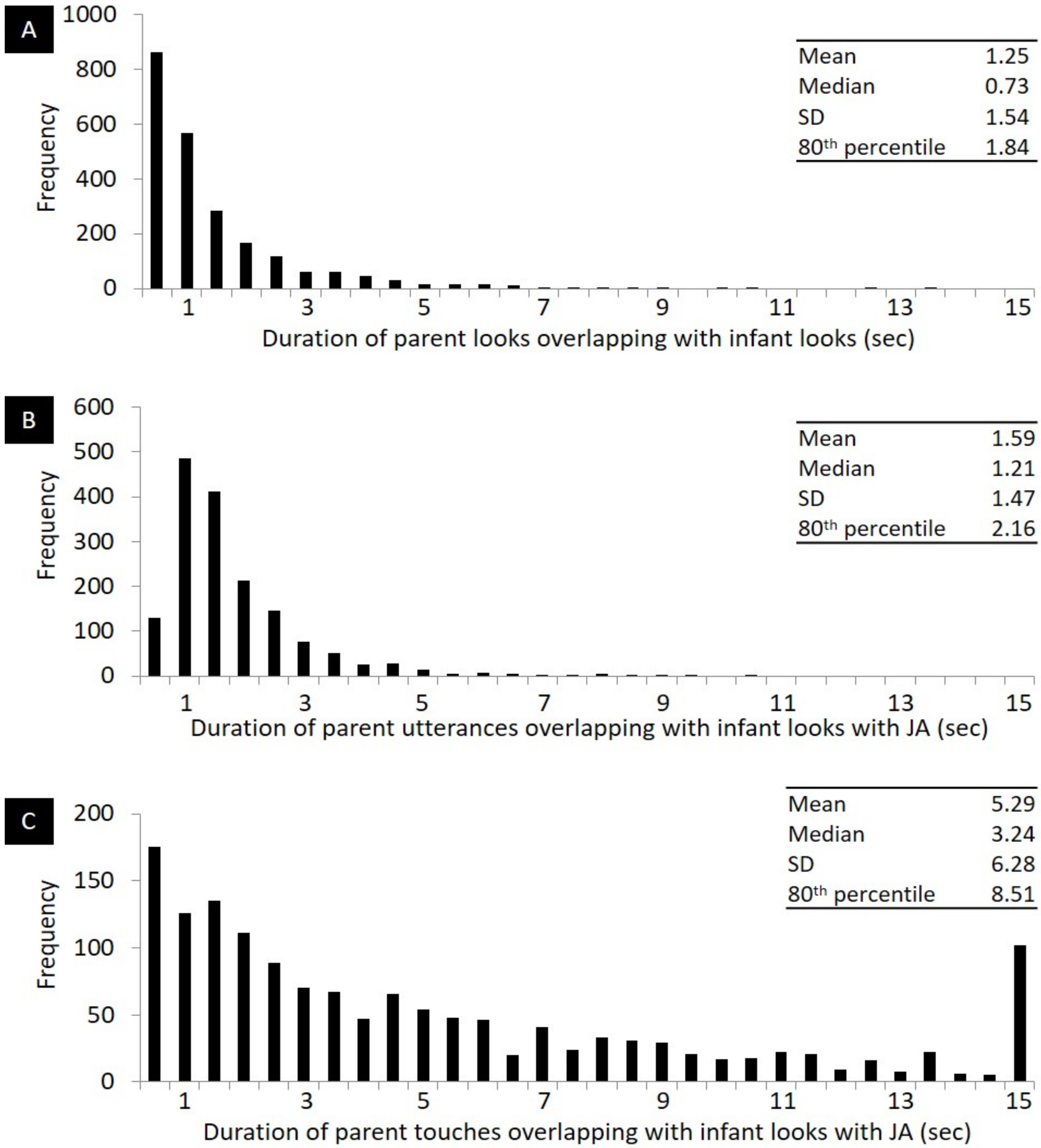


Figure 4. Histograms of the duration of parent behaviors (A: parent looks, B: parent utterances, C: parent touches) overlapping with infant looks, and descriptive statistics of these distributions. Note the Y axes are on different scales reflecting the different properties of parent looks, talk and touches to objects.

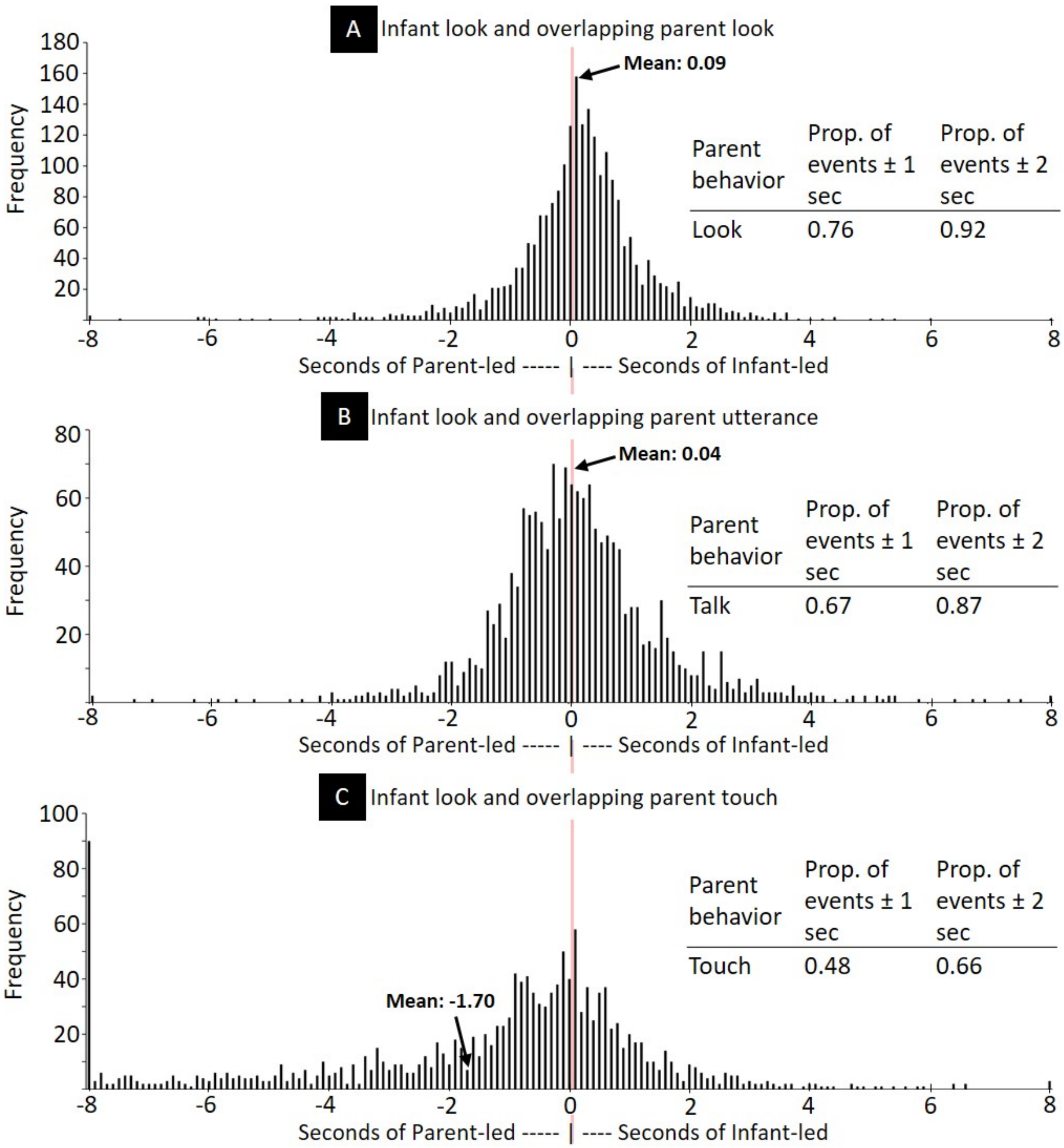


Figure 5. Histogram showing lags in seconds between onset of first parent's look, utterance, touch (shown in A, B and C respectively) and the onset of the infant's look. The line at zero shows all panels are aligned and it represents moments in which the onsets of infant's look and the parent's behavior occurred simultaneously with lag=0. Note the Y axes are on different scales reflecting the different dynamic properties of looks, talk and touches to objects.

Figure 6 compares the durations of infant looks that include coordinated joint parent attention to the object to those that do not. As it is evident, the frequency of infant looks with JA was greater, 9.88 per min (SE=0.33) than those without with Coordinated JA, 5.88 per min (SE=0.35). More critically, the duration of infant's individual looks were longer with JA than without. That is, we observed the same pattern reported by Yu & Smith (2016): the duration of infant looks to an object was longer when parents also looked at the object (M=3.03sec, SD=0.69, Median=2.95, SE=0.11) than when they did not (M=1.26sec, SD=0.26, Median=1.18, SE=0.04), t(39)=17.64, *p*= 2.2e-16). We also performed a Wilcoxon Signed-rank test because whereas the difference scores used in the t-test are normally distributed, the durations of infant looks with and without Coordinated JA themselves are not. The Wilcoxon Signed-rank test also confirmed that infant looks with Coordinated Joint Attention were longer in duration than infant looks without Coordinated Joint Attention, Z=-5.51, *p*<0.001. Figure 6 suggests that two distributions of infant looks –with and without Coordinated JA –differ primarily in the longer tail of long looks. Accordingly, we determined the normalized count of each infant's Long looks with Coordinated JA and without Coordinated JA, the previously used threshold for sustained attention (Yu & Smith, 2016; Ruff & Lawson, 1990), dividing the count of Long Looks with Coordinated JA, for each subject, by that subject's total number all looks with Coordinated JA and likewise, the count of Long Looks without Coordinated JA by total number of looks without Coordinated JA. There were more Long looks with JA (Mean proportion=0.37, SD=0.10, Median=0.36. SE=0.02) than without JA (Mean proportion=0.05, SD=0.05, Median=0.05, SE=0.01), t(39)= 19.01, *p*=2.2e-16). The Wilcoxon Signed-rank test confirmed this finding, Z=-5.51, *p*<0.001. Very Long looks –at least 10-seconds-long— by the child, only occurred when parents also shared attention to that object (Mean proportion=0.03, SD=0.03, Median=0.02, SE= 0.005). The duration of infant looks with Coordinated JA did not differ as a function of who led the look to the object (parent-led: 2.91sec, SD=0.94, Median=2.67, SE=0.15; infant-led: M= 3.05sec, SD=0.70, Median=2.97, SE=0.11; $t(39)=1.03$, $p=0.31$, ns. A Wilcoxon Signed-rank test confirmed this finding, $Z=-1.18$, $p=0.24$, ns. In brief, the main finding is this: There were more Long unbroken looks to a single object by infants when their parent also looked to the same object during the infants' unbroken look.

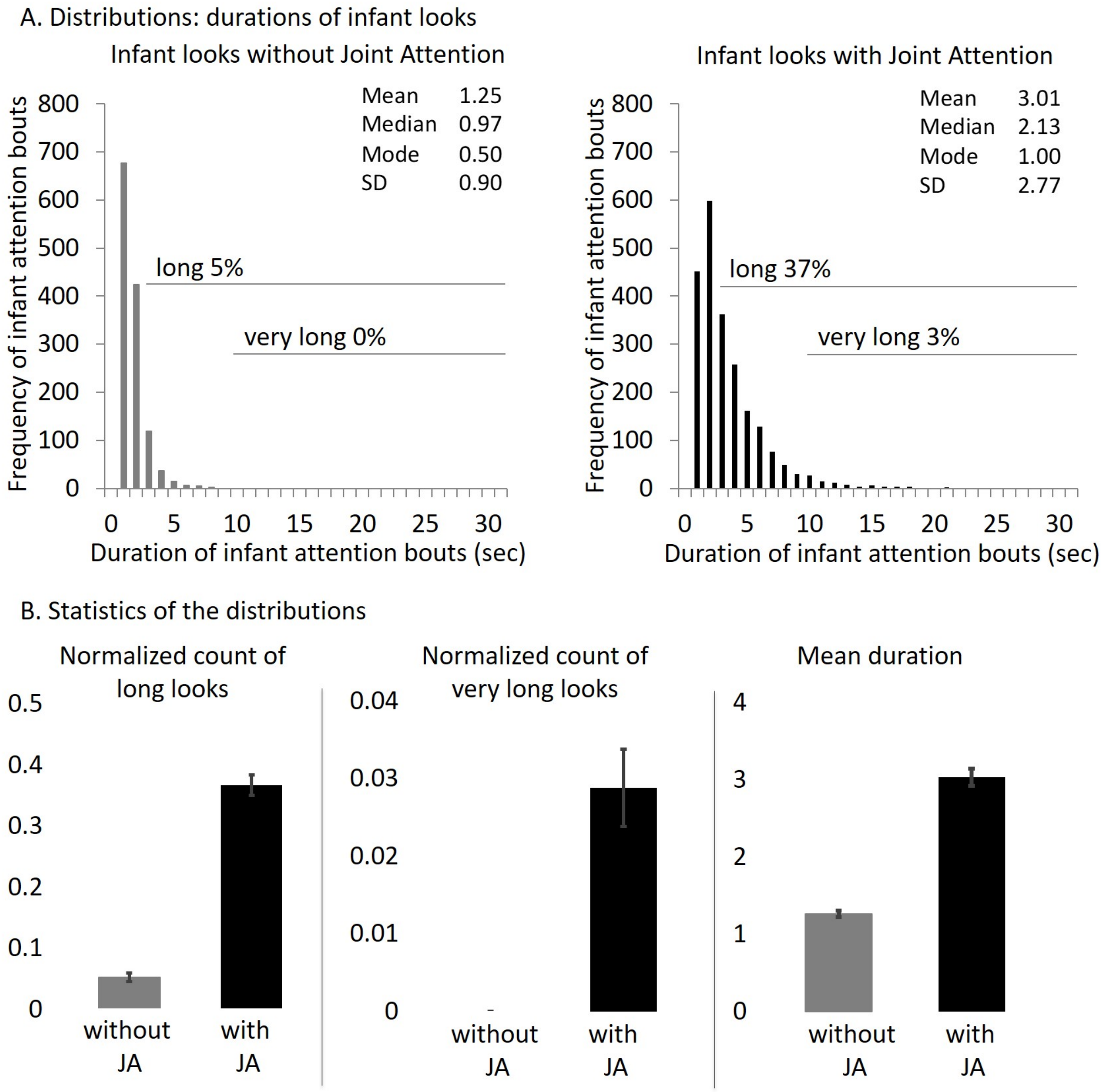


Figure 6. A. Histograms of duration of infant looks without and with Coordinated JA. B. Three different statistics of each distribution illustrated in A are compared: Normalized count defined as the mean proportion of each infant's looks that were Long (>3sec) and Very Long (>10sec) given that the infant look did not or did include Joint Attention. We also compared the mean duration of the distributions of infant looks without and with JA computed across subjects. Error bars in this graph represent standard errors around each of the means.

Because infant looks to the object are defined as unbroken looks, with no glances away from the object, the infant –when looking at the object– cannot be looking directly at the parent's face. Given this, it is unlikely that the infant during their sustained looks to an object used the parent's look itself as the indicator that the parent was simultaneously attending to the same object. In addition, peripheral vision seems an unlikely source of such information as considerable evidence indicates that gaze following is very difficult for adults, older children and infants in natural contexts with freely moving heads and objects; indeed, success in following gaze —for adults, children and infants — is quite poor in any context in which there are not just two choice objects widely separated in space on opposites sides of the midline of the person directing gaze (Corkum & Moore, 1998; Doherty, Anderson & Howieson, 2009; Farroni, Johnson, Brockbank & Simion, 2000; Langton, Watt & Bruce, 2000; Loomis, Kelly, Pusch, Bailenson & Beall,2008 ; Vida & Maurer, 2012a-2012c). In principle, it is also possible that infants looked at the parent's face *just prior* to looking at the object themselves and registered a prior look to the object by the parent. This also does not seem likely since parents follow infants' looks to an object more than they lead (Figure 5A). Although we cannot definitively rule out that infants had some direct knowledge that their parent looked at the object by directly perceiving the parent's look while the infant was looking at the object, it seems more likely that other parent behaviors that overlap with infant and parent looking at the object may be the behaviors that influenced the infant's continued visual interest in the object and signaled parent's engagement with the object.

III. Multimodal Parent Behaviors

Given that an infant is looking at an object and the parent also looks to that object, what else does the parent do that might influence the duration of infant attention? Figure 7 illustrates how we identified other potentially relevant combinations of parent behaviors —talk and touch of the infant-attended object— for consideration. These additional behaviors were located as potential influences if they overlapped with *an infant's look* that was part of a joint attention episode. The additional parent behavior could be long, or short in duration and could or could not overlap in time with the parent look that defined the infant's look to the object as including joint attention. The additional parent behavior also could precede or follow the parent look to the infant attended object. That is, , the *only* —and objective —criterion for considering the additional parent behavior's effect on infant looking was that the parent behavior overlapped in time with the infant's look to the *jointly* attended object. In this way, we know the parent was attentive to the object of infant interest (not some other object) and thus that these additional behaviors likely reflected that interest.

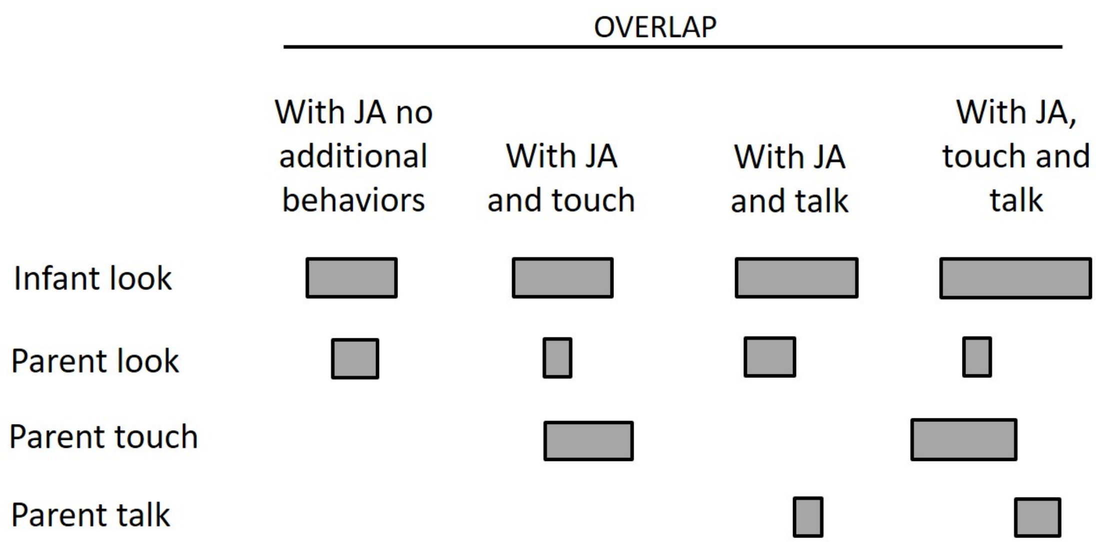
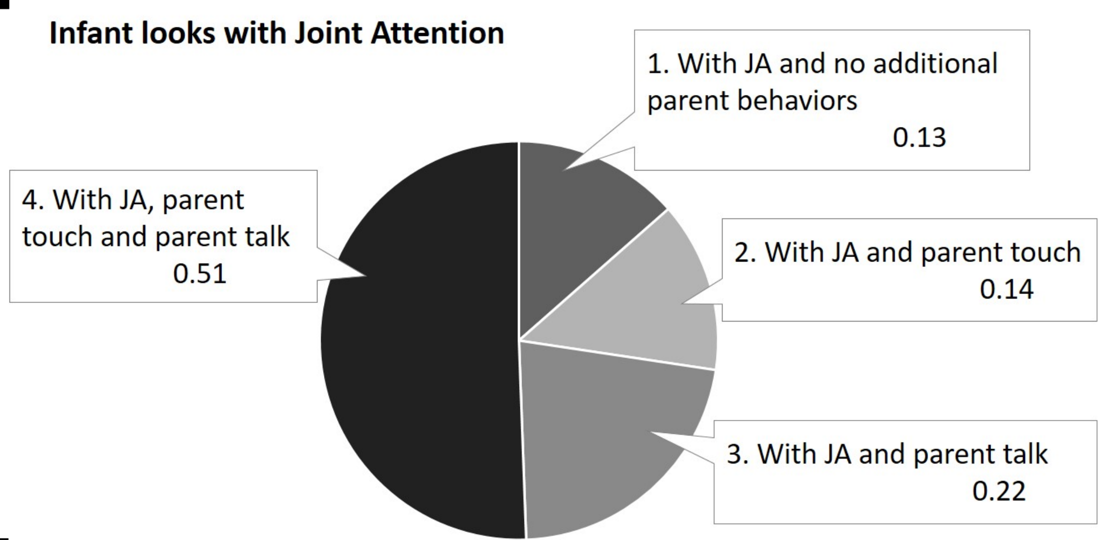


Figure 7. Definitions for infant looks *that included Joint Attention* and different combinations of parent behaviors. Overlap with a parent behavior (touch or talk) was defined objectively as the temporal overlap between an infant look to an object and the parent behavior. The possible combinations of parent behaviors yielded four categories of infant looks that are also referred as

JA categories, as they all have an overlap between the infant look and the parent look to the infant-attended object.

Figure 8A shows the proportion of *all infant looks* with JA that overlapped in time, as described, with the four possible combinations of the additional parent behaviors and, as expected, it indicates that there was much more going on than just parent and infant looking the same object. When the infant was looking at an object, and the parent also looked at the object, parents added at least one additional behavior — talk or touch— as an indicator of their interest, over 87% of all JA episodes. This fact indicates that joint attention typically occurred with more multimodal parent *engagement* with the object. Infant look durations for these four categories of infant looks with JA along with infant looks without JA are the data submitted to the main analyses. Accordingly, Figure 8B shows the mean proportions of the count of *all* infant looks (with and without JA) that were of each of the four JA categories or were without JA. Figures 4B and 4C show the histogram of the durations of parent talk and hand contact with to the infant- and parent-attended object. Both distributions are again extremely skewed with most events being quite brief, but talk is overall much briefer and consistent in its timing whereas some touches can be quite long and are more variable (note 80th percentile of utterance duration is much smaller than the 80th percentile of touch duration). Figures 5B and 5C show the timing of parent talk and touches relative to the onset of the infant look. As is the case for parent looks (Figure 5A), parent talk and parent touches are all centered on the onset of the infant look to the object and because of the tight coupling of parent look onset to infant look onset, they are also centered on the onset of the parent look. Figures 5B and 5C show the proportion of events that fall within 1 second and within 2 seconds of onset of infant look for both talk and touch respectively. These results show the onset of parent touch is less coupled in time with the onset of the infant's look as compared to the onset of parent look and talk. More than two thirds of the data are within -1 and 1 seconds for both parent look and parent talk but not for parent touch. The key question is the potential influence of these multimodal parent behaviors on the duration of infant attention to the object.

A



B

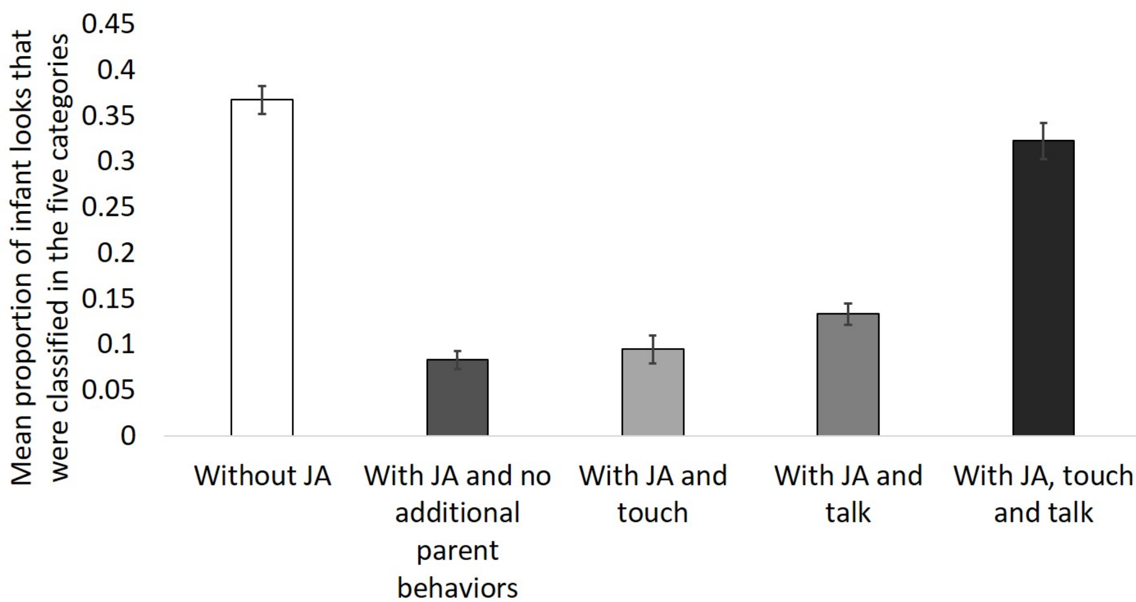


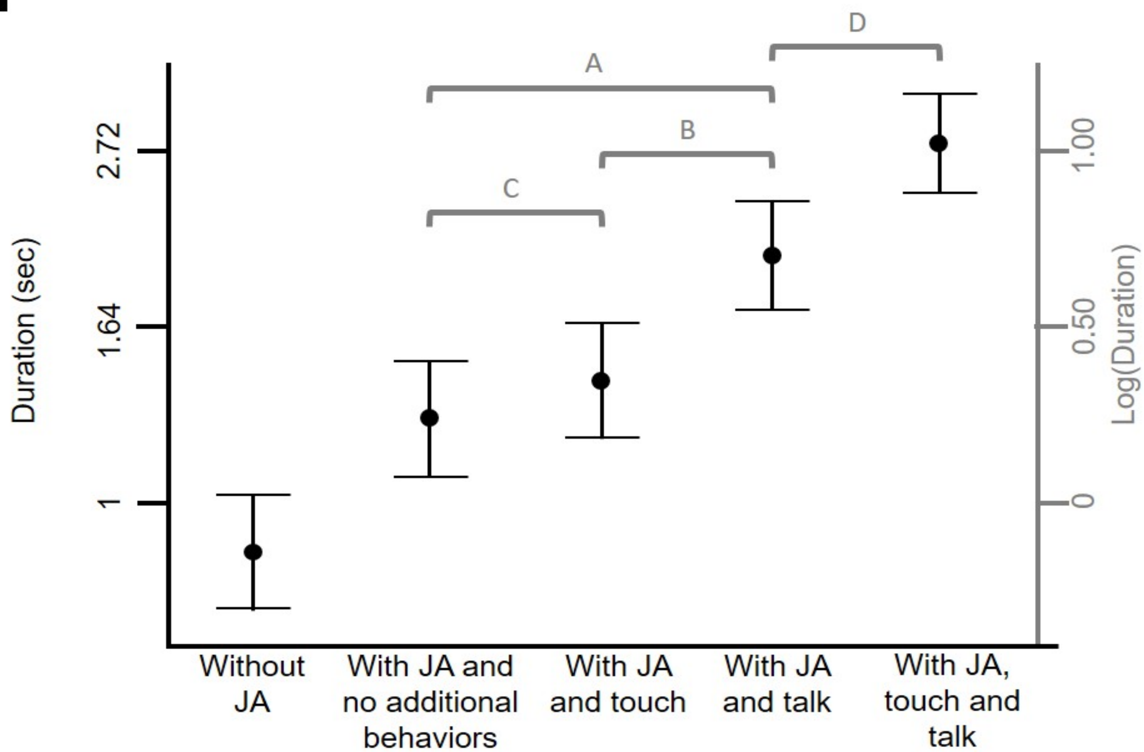
Figure 8. A. Proportion of all *infant looks with JA* that were classified as each of the four categories of JA: JA with no additional parent behaviors, JA with parent touch, JA with parent talk, and JA with parent touch and talk. B. Mean proportion of *all infant looks* (with and without JA) that were classified as each of the five categories entered into the main analyses. The error bars in this graph represent standard errors around each of the means.

The observed often tight temporal coordination of parent looks, talk and touch of the infant-attended object emphasizes the complexity of naturalistic free-flowing parent behavior and the difficulties in singling out the role played by any individual component of parent behavior. Accordingly, the approach we took is to define four categories of parent behavior during infant attention to an object, given that the parent had also looked to that object (also shown in Figure 7): (1) the parent looked at the infant attended object but did not talk or touch; (2) the parent looked and also touched the infant attended object, (3) the parent looked at the infant-attended object and also talked, and (4) those in which parents looked at the infant-attended object, talked and touched it. We compare the durations of infant looks in these four categories with a baseline of the duration of infant looks when parent did not also look at the object, yielding 5 categories in the analysis. Because durations are not normally distributed but skewed, we submitted the logs of the durations of the infant looks to a linear mixed-effects models (code online @ *https://github.com/csuarezr26/MultimodalJA_andSA2018*) using the lme4 package version 1.1.12 (Bates, Maechler, Bolker & Walker, 2015), and the lsmeans package version 2.27.2 (Lenth, 2016) in the R environment (version 3.3.2) (R Development Core Team, 2016). The model predicted the log(duration) of individual episodes of infant attention from a fixed effect, which specified the 5 possible categories. Random intercepts were specified for individual infants -which controls for infants having different durations of attention bouts- and for the specific object attended to by the infant –which controls for different objects having different durations of attention bouts- (i.e., random intercept model, Pinheiro & Bates, 2000; for an application, see Oberauer & Kliegl, 2006). Finally, the package used restricted maximum likelihood specifying an unstructured covariance matrix. We used this linear mixed model because it allowed us to ask how the specific combinations of parent behaviors may differentially predict the duration of attention bouts while accounting for the fact that duration of infant looks are not independent of each other and could vary across infants.

Figure 9A illustrates the results, showing the predicted mean duration and 95% confidence interval of infant looks in the 5 categories specified in the model. First, the durations of infant looks given that a parent looked at an object, the four JA categories, were each significantly longer than infant looks to an object without a parent look to that same object. JA episodes that included both parent talk and touch coincided with the longest looks by the infant to the object. Table 2 shows the estimated difference, SE and p-values with respect baseline category for each of the four categories of JA. The model's intercept, in logarithmic scale, was -0.14, which corresponds to an estimated mean duration of 0.87 seconds (as the natural exponential of -0.14 is 0.87, exp(-0.14)=0.87) for the duration of the baseline category (infant looks without JA). The estimated differences shown in Table 2 are in logarithmic scale and they represent the difference between the predicted mean log(durations) of each of the categories and the mean log(duration) of baseline (0.87 seconds, also shown in Figure 9A). The values in Table 2 can be converted, by computing their natural exponential, to represent also the increase in mean duration from baseline, for each category, as a factor of the duration of the baseline category. In this way, for instance, the estimated mean duration of category 2 (infant looks with

JA and no additional behaviors) is 1.46 times the duration of the baseline category (as exp(0.38), the natural exponential of 0.38, is 1.46). The estimated mean duration of category 4 (infant looks with JA, touch and talk) was 3.19 times the duration of the baseline category (exp(1.16)=3.19)).

A



B

	Contrast		Estimated difference	SE	p-value
A	With JA and no additional behaviors	With JA and talk	-0.46	0.05	<.0001
B	With JA and touch	With JA and talk	-0.35	0.05	<.0001
C	With JA and no additional behaviors	With JA and touch	-0.11	0.06	0.3455
D	With JA and talk	With JA, touch and talk	-0.32	0.04	<.0001

Figure 9. A. Estimated mean log(duration) of infant looks in the five categories of infant looks with 95% confidence intervals around the estimated means. Letters A-D illustrate the contrasts tested in the model. B. Beta coefficients and standard errors of the planned contrasts A-D. The p-

values are adjusted to account for all possible 5X5 pairwise comparisons according to the Tukey correction.

Table 2

Beta Coefficients and Standard Errors of the Linear Mixed Model

Contrast		Estimated difference	SE	p-value
2: with JA no additional behaviors	1: without JA	0.38	0.05	<.0001
3: with JA and touch	1: without JA	0.49	0.05	<.0001
4: with JA and talk	1: without JA	0.84	0.05	<.0001
5: with JA, touch and talk	1: without JA	1.16	0.04	<.0001

As is evident, parent talk had a much greater effect on the duration of infant looks to the object than did touch. Pairwise contrasts (Figure 9B) show: (A) Infant look durations with JA that included parent talk were longer than infant look durations with JA with no additional parent behaviors; (B) Infant looks with JA that included only parent talk were longer than those with JA than included only parent touch of the object; (C) Infant look durations overlapping with JA and parent hand contact were not reliably longer than those that included no additional measured parent behaviors; (D) However, JA episodes that included both parent talk and parent hand contact with the objects were associated with longer infant looks to the object compared to JA episodes including parent talk alone, indicating that the additional behavior of touching the attended object does add to the strength of parent interest expressed by talk. The fit of the overall model decreases if the single fixed factor (the five categories based on parent behavior) is removed, leaving only the random factors as predictors: $\chi^2(4) = 715.58$, $p < .00001$. We compared the model to a linear mixed effect model that included random slopes for infants' effect of category of look on log(duration) of looks in order to test if the more complicated model, with random slopes, provided a better fit. The results suggests this was not the case as both models provided a similar fit for the data ($\chi^2(14) = 18.48$, $p=0.18$).

One open question raised by the results in Figure 9A is the role of parent looks to the objects —either with or without the additional behaviors. Because infants' rarely look directly to the parent face during the session and do not do so during the measured dependent variable of continuous gaze directed to an object, it seems unlikely that parent looks in and of themselves directly influence the duration of the infant look to the object. However, the results show that a parent look to the object is a critical component of multimodal parent behavior.

First, parent talk without an overlapping parent look was not associated with more enduring infant looks to the attended object. Infant looks that overlapped with talk but did not include joint attention (or an overlapping parent look to the same infant-attended object) were much shorter (M= 1.41 sec, SD=0.34, Median=1.34, SE=0.05) than infant looks that included talk and also joint attention (M=3.48 sec, SD=0.90, Median=3.28, SE=0.14), $t(38)=15.46$, $p<0.001$. A Wilcoxon Signed-rank test confirmed that infant looks *with* joint attention and with talk were longer in duration than infant looks *without* joint attention and with talk, Z=-5.44, $p<0.001$. Second, parent touches without an overlapping parent look as well, were not associated with longer lasting infant looks to the objects. Infant looks that overlapped with touch but did not include joint attention were significantly shorter (M= 1.30 sec, SD=0.39, Median=1.18, SE=0.06) than infant looks that overlapped with touch and also with joint attention (M=3.39 sec, SD=0.78, Median=3.29, SE=0.12), $t(39)=17.05$, $p<0.001$. A Wilcoxon Signed-rank test confirmed this finding, Z=-5.51, $p<0.001$. These results suggest parent talk and touch without a coordinated parent look to the infant-attended object have more limited influences on infant's attention compared with multimodal parent behaviors that also include parent looks.

One limitation on these conclusions is the rarity of talking or touching by a parent when the parent is not also visually attending to the object attended to by the infant. Infant looks that overlapped with talk but not with a parent look occurred on average 3.07 times per minute of eye-tracking data (SD=1.46, SE=0.23) while infant looks that overlapped talk with JA occurred on average 9.37 times per minute of eye-tracking data (SD=2.30, SE=0.37). Infant looks that overlapped with touch but not with a parent look occurred on average 0.84 times per minute of eye-tracking data (SD=0.73, SE=0.12) while infant looks that overlapped with touch with JA occurred on average 8.37 times per minute of eye-tracking data (SD=2.15, SE=0.34). This is not surprising; human behavior is a coordinated multimodal event in which we look at what we handle (Yu & Smith, 2013) and look at the referents of our talk (Henderson & Ferreira, 2004; Griffin & Bock, 2000).

However, parents sometimes looked at the attended object without also touching it or talking and these moments also were associated with longer infant looks to the object than when parents did not also look at the object (Table 2). If infants cannot directly perceive parent's eye gaze in this circumstance, then parents must be showing their interest through other unmeasured behaviors, perhaps leaning forward or hands close to but not touching the object, or perhaps even a stillness of body so as not to disrupt (and/or redirect) the infant's attention. This is a question for future research.

In summary, the main finding appears to be this: During joint parent-infant interaction with objects, parents often look at the objects to which their infants are visually attending and when they do, they often talk and touch the object during the period of the infant's attention to that object. All of these behaviors appear to increase the duration of the infant's attention to the object. The most potent combination —which also happened most often in free-flowing interaction— is a multimodal behavior that includes the parent's look, talk and touch and increases the duration of infant look to the object by about 1.5 sec (on average) as to when the parent displays no interest at all. This may seem a relatively small increase; however, as proposed by Yu and Smith (2016) and as we consider in the General Discussion, these small increases —repeated over and over within a single play session and across the days and weeks of the infant's life— may have long term consequences on infant development.

Discussion

The results show that infant sustained attention is more likely and more enduring when parents also visually attend and express interest through talking or handling the object attended to by the infant, providing converging evidence for a role of social context in infant sustained attention (Yu & Smith, 2016). The new contributions are that in addition to parent look, multimodal behaviors are embedded within these joint attention moments and that both parent hand actions and parent talk within the context of shared attention to an object sustain infant visual attention to the jointly attended object. The findings have implications for current understanding of the relation between joint attention and sustained attention, and, most critically, for experiential factors in the development of sustained attention.

Joint Attention and Sustained Attention

Joint attention and sustained attention have been studied separately but not together. Joint attention has been traditionally examined in social contexts, with parent gaze understood as a potential social signal to be read and responded to by the infant (e.g, Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005); infant sustained attention has been considered as an individual achievement in early development (Colombo, 2001; Pérez-Edgar, McDermott, Korelitz, Degnan, Curby, Pine & Fox, 2010; Posner & Rothbart, 2000; Ruff, 1990). But parent-infant joint attention involves more than gaze behavior and contains the infant's own visual attention to the object. Parents do not just look to objects when they share attention with their infant; they express their interest in multimodal behaviors, which have direct effects on infant visual attention. Thus, joint attention may be best conceptualized not as shared visual attention to an object but as a proxy of a suite of multimodal temporally coordinated parent and infant behaviors. The present results clearly show the fine-grained temporal coordination of the onsets of parent multimodal behaviors directed to an object –gaze, talk, touch– and the onset of infant visual attention. The lag between the onset of any of the three types of parent behavior and the onset of infant looking fell at near zero in all cases (shown in Figure 5). This tight temporal entrainment of infant looking and multimodal parent behaviors marks joint attention episodes as a potentially powerful context for multiple aspects of human development, including the self-regulation of attention. One implication is that the strong predictive relations between joint attention and other developments such as word learning (Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007; Tomasello & Farrar, 1986;) may emerge not exclusively through shared gaze but through the other parent behaviors that are part of naturally occurring episodes of parent-infant joint attention. This idea is further supported by a recent study (Yu, Suanda, & Smith, in press), showing that infant sustained attention but not joint attention in toy play is predictive of later language outcome.

The multimodal nature of joint attention as evidenced by the parents in this study may help understand how joint attention works in the wild, given that infants engaging in object play rarely look to the faces of their partners (Bakeman & Adamson, 1984; Deak, Krasno, Triesch, Lewis & Sepeta, 2014; Franchak, Kretch, Soska & Adolph, 2011; Yoshida & Smith, 2008; Yu & Smith, 2013). The operational definition of joint attention used in the present study differs from the definition of "joint attention" used in many discrete trial experimental studies. (e.g, Baron-Cohen & Cross, 1992; Brooks & Meltzoff, 2005; Mundy, Block, Delgado, Pomares, Van Hecke & Parlade, 2007). In those studies, researchers often sought evidence that the child was aware of the mature partner's direction of attention, requiring infant looks to the partner's face for coordinated visual attention to the same object to count as joint attention. Here, we used a simpler and more objective measure, the degree to which parent and child directed gaze to the same object at the same time, a measure we believe is more fitting for naturalistic object play (Yu & Smith, 2013; Yu & Smith, 2017). However, just because the infant does not look at the parent face does not mean that the child is unaware of the parent's direction of attention. Hands, talk, as well as other yet unmeasured behaviors are likely well-read cues by infants about their parent's current state of interest. The close temporal coordination of parent looks, talk, and touch have been noted by other researchers as a powerful combination that makes parents referential intentions transparent (Trueswell, Lin, Armstrong, Cartmill, Goldin-Meadow & Gleitman, 2016) and indicates timely and coordinated responsivity on the parents part to infant interests (Van Egeren, Barratt & Roach, 2001). Thus, instead of focusing solely on parent gaze and child gaze in joint attention, the present study suggests a broader context to examine joint attention embedded in the multimodal parent behaviors that take place in naturalistic social interactions.

Moreover, the results also suggest the value of a more unified study of joint attention, infant sustained attention, and parental responsiveness. First, if episodes of joint attention are the context in which parents scaffold the development of infant sustained attention and if that scaffolding requires rapid responses from parents to the infant's attentional state in fractions of seconds, then the complete explanation of the development of sustained attention –and individual differences in that development– will depend on understanding how all of these components interact and how they vary across dyads. Second, numerous studies directed to these components –joint attention, infant sustained attention, parent responsiveness– show they each predict later outcomes. The present results raise the possibility that joint attention and parental responsiveness are predictive because they support the development of sustained attention and the self-regulation of attention that is essential to learning in all domains. This hypothesis can be correct only if parent behaviors have their effects on infant attention through processes related to the development of the self-regulation of attention.

A Training Ground for the Self-Regulation of Attention?

The importance of parent behavior in supporting infant attention to an object certainly lies in the in-the-moment effects that this behavior may have on the duration of the infant's attention and thus on infant learning in the moment. A recent study showed that infants engage in sustained attention more in the context of parent-infant joint play and less during contexts in which the infant plays with objects alone (Wass, Clackson, Georgieva, Brightman, Nutbrown & Leong, 2018a). The accumulating evidence points to the implication that mature social partners may support in-the-moment visual learning of the attended object –documented by studies that examine infant sustained attention alone (e.g., Ruff, 1986)—through their effect on the infant's visual attention. However, the more intriguing possibility is that these day-in day-out small effects of parent behavior on sustained attention to an object may serve as the experiential training ground for longer term effects. How might such a mechanism work? There are multiple inter-related mechanisms that may be involved. For example, parent behavior could signal parent interest, and the infant's awareness of that interest may be a factor that keeps the infant attending to the object. It also possible that parent expressed interest may be rewarding to the infant, engaging reward mechanisms that have been shown in adults to support persistence in following goals and avoiding (Insel, 2003; Montague, Hyman & Cohen, 2004). Words are well known drivers on visual attention in infants as well as older children (Carvalho, Vales, Fausey & Smith, 2018; Fernald, Thorpe & Marchman, 2010; Johnson, McQueen & Huettig, 2011; Landau, Smith & Jones, 1992; Vales & Smith, 2015;), and the most powerful parent behavior observed in the present study. Finally, the dynamic coordination of parent and infant behavior in real time may operate in real time more like a dyad dancing together to socially entrain (and thus train) internal attentional control mechanisms (Marsh, Richardson & Schmidt, 2009; Takahashi, Narayanan & Ghazanfar, 2013). There is some evidence consistent with the proposal that increased exogenous (externally driven) attentional capture that is likely created by the parent is responsible for the increase in infant's sustained attention appearing in contexts of joint play (Wass et al., 2018a; Wass, Noreika, Georgieva, Clackson, Brightman, Nutbrown, Covarrubias & Leong, 2018b). The distinction between top-down versus bottom-up processes as well as between endogenous versus exogenous drivers of attention during joint play with a social partner are key issues for future research.

In sum, the present findings strongly implicate that multimodal parent behaviors during joint attention may –in one way or another– influence sustained attention in real time, with the potential to influence the internal mechanisms that underlie the development of the self-regulation of attention. The effects of any single joint attention bout may be quite small with the infants' gaze to the object extended only slightly (Yu and Smith, 2016). Nonetheless, the aggregated effects on the development of self-regulated attention may be quite large as these small socially guided extensions of infant visual attention may occur multiple times a day, day-in and day-out in the social lives of infants. The next chapter of this dissertation examines the relation between these moment-to-moment effect and more long-term effects, testing the possibility that Coordinated Joint Attention supports the development of infant's own attentional control.

Chapter 4

Socially Scaffolded Sustained Attention Predicts Future Independent Sustained Visual Attention

The self-regulation of attention is a key developmental achievement that supports success in school and in social interactions. One early marker of self-regulated attention in late infancy is the ability to sustain attention on one toy when playing with multiple available toys (Ruff, 1990; Ruff & Lawson, 1990). Within this context of self-generated play, one-year-old infants rarely sustain visual attention to an object whereas two year old children are much more likely to punctuate rapid shifts of gaze among available toys with periods of sustained visual attention to an individual toy. Because early individual differences in sustained attention in toy play predict later individual differences in assessments of self control in older children, sustained attention during toy play and self-controlled attention in childhood are believed to reflect the same underlying system of endogenous attentional control (Ruff, 1990; Ruff & Lawson, 1990; Ruff, Lawson, Parrinello & Weissberg, 1990). Several recent studies have shown that although one-year old infants are not likely to sustain attention on a single object when playing alone, they are much more likely to do so when playing in the company of a parent. Parents through their behaviors –looking at, talking about touching the infant-attended object (Chapter 3) – appear to actively encourage their infants to stay engaged with a single object for several seconds longer than they would on their own (Yu & Smith, 2016).

What is the role of these early parent-supported moments of sustained attention in the development of *self*-controlled sustained attention? Socially scaffolded sustained attention could be a critical part of the developmental pathway to self-regulation (Chapter 3 and Yu & Smith, 2016) or a quite different phenomena (Wass et al., 2018a). Here we provide relevant evidence showing that effective parent-scaffolding of long looks to an object at 16 months predicts the sustained attention of infants when playing alone at 24 months. We show further that the frequency of infant sustained at 16 months without strong support from parent behaviors does not predict self-generated sustained attention at 24 months.

Figure 1 illustrates the experimental approach. At 16 months, we measured the duration of infants' *individual* looks to an object (see Figure 1) when parents scaffolded individual looks with three overlapping behaviors, looking at, talking about, *and* touching the object. We defined Scaffolded looks as requiring all three overlapping parent behaviors because prior work showed that when parents jointly attended to an infant-attended object they usually also talked about and touched the object, and that individual infant looks that overlapped with all three parent behaviors (looking, talking, touching) were markedly longer than when parents were less engaged. At 24 months, the children were given a set of toys to explore on their own and the duration of their gaze to individual objects was measured.

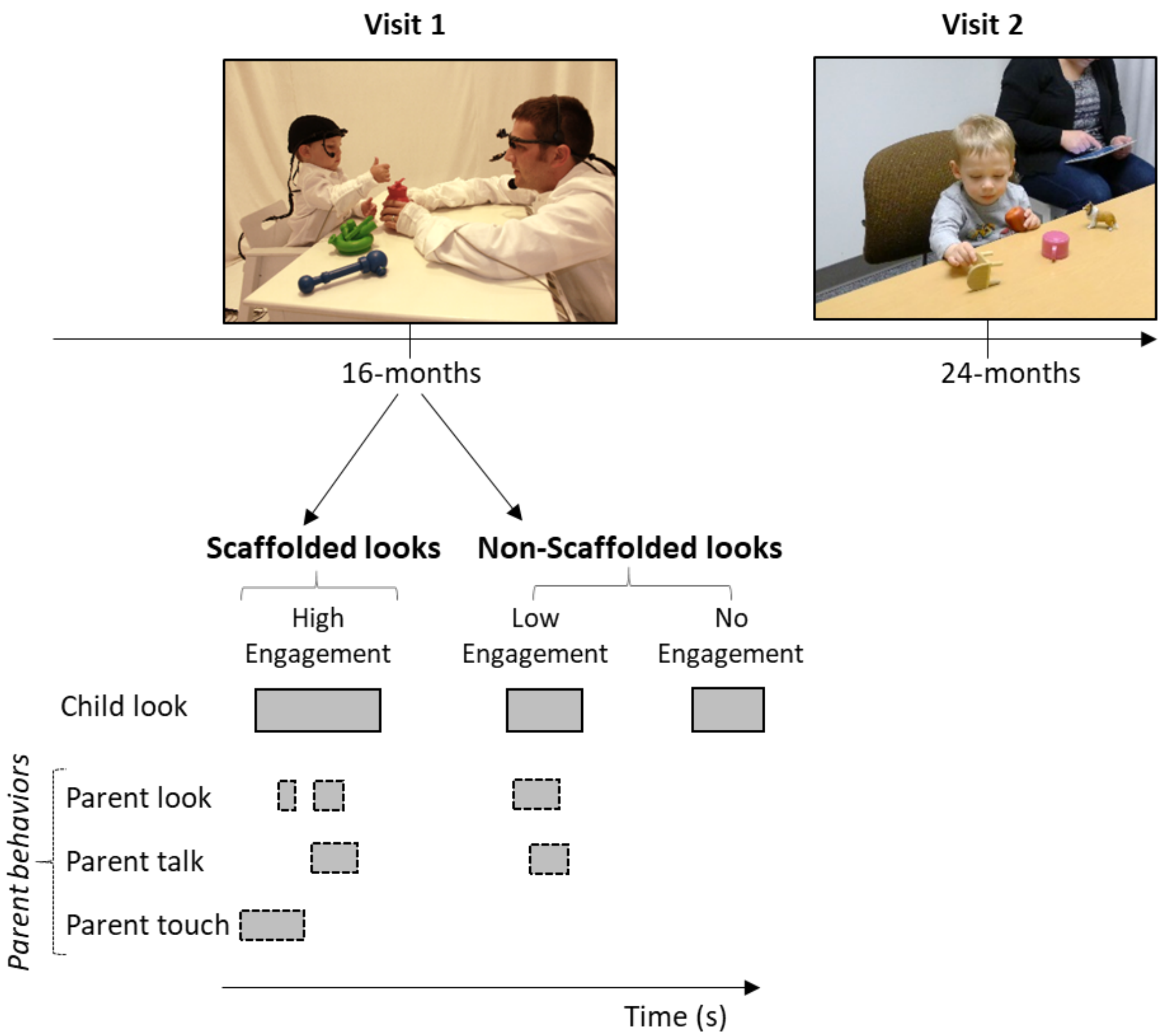


Figure 1. Visit 1 at 16-months consisted of parent-infant play, while visit 2 at 24-months consisted of object exploration by the child alone. Measures of child eye-gaze were collected during both visits and infant looks were defined as unbroken gaze to a single object. Infant looks at 16 months were classified as *Scaffolded Looks* or *Non-Scaffolded looks* based on the temporal overlap of each infant's look with the three parent behaviors (look, talk and touch).

Method

Participants

The participants were 28 children (12 females) and their parents, who were also contributing data to a larger longitudinal project. For ease, we will call the children infants at the 16 month visit and children at the 24 month visit. At the 16-months visit, the infants were on average 16.46 months (SD=1.49). The age selected for the second visit was 24-months, an age at which self-directed and sustained object exploration appears in some children (Ruff & Lawson, 1990). At the 24-months visit, children were on average 24.48 months (SD=0.66). All 28 children contributed data during both visits. Seven other infants were recruited but did not contribute data because they refused to wear the head-mounted device or other technical failure. The entire sample of children was broadly representative of Monroe County, Indiana (84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Other) consisting of predominantly working- and middle-class families. This research project was approved by the Institutional Review Board at Indiana University (protocol number 0808000094) and titled “Multimodal word learning”. Participating children were given a small toy as compensation.

Stimuli

Twelve novel objects (on average, about 9.50 cm ^{x} 6.5 cm ^{x} 5.0 cm), each with a unique shape and color, were designed for object play at the 16-months visit. All objects in the pool were piloted and tested to be visually and manipulatively engaging. Each infant only played with 6 objects that were selected from the pool of objects to not overlap with any previous objects used in the longitudinal study. Infants played with the six objects in 2 sets of three, and in each set one object was painted blue, one red and one green. The stimuli for the 24-month visit consisted of four familiar toys, designed to elicit exploration and play by the child. The toys included a cup, an apple, a dog and a chair. These objects were on average 6.5 cm ^{x} 6.25 cm ^{x} 5.12 cm.

Experimental Room

At the 16-months visit, the parent and infant sat at a small table (dimensions 61 cm ^{x} 91 cm ^{x} 64 cm) facing each other. The center of the tabletop was approximately 46 cm away from the infant's eyes and the parent's eyes. The infant sat on a chair and the parent sat on the floor. Participating parents reported this posture to be natural and comfortable. Both participants wore head-mounted eye trackers (Positive Science LLC, http://www.positivescience.com; also see Franchak & Adolph, 2010) during the play session. The eye-tracking system attached to the infant's head through Velcro sewed into a hat and thus remained stable throughout the experiment. The eye-tracking systems for both the infant and parent included an infrared camera that recorded eye images –mounted on the head and pointed to the right eye of the participant– and a scene camera that captured the events from the participant's egocentric perspective. The scene camera's visual field was 108 degrees, capturing a broad view although less than the full visual field. Each eye tracking system recorded both the egocentric-view video and eye-in-head position (x and y) in the captured scene at a sampling rate of 30 Hz. In order to allow for automatic detection of the objects on the table, everything in the room –other than the objects and the hands and faces of the participants- was white. This environment elicited natural play centered around the objects on the table by the participating dyads. Three additional cameras recorded the interaction from third-person views.

At the 24-months visit, the child sat in front of the parent at a table (dimensions 151 cm ^{x} 76 cm ^{x} 66 cm) facing an experimenter. This position encouraged children to explore the objects presented by the experimenter independently without the parents' engagement. The experimenter sitting in front was approximately 75 cm from the center of the table and the child was 63 cm. In order to encourage children's independent play and to limit testing time (as this test was included during a testing session for the larger longitudinal study), the child did not wear a head-mounted eye-tracker. Instead, a camera was directed at the center of the child's face and gaze direction was coded frame-by-frame from this recording, a reliable procedure widely used in many studies (Miller et al., 2009). Two additional videorecorders also recorded the play – an overhead birds-eye view of the whole event, and a side-view of the interaction.

Procedure

Upon entering the experimental room at the 16-months visit, a second experimenter and the parent showed a highly interactive pop-up toy to the infant as the first experimenter placed the head gear on the infant. Infants were previously desensitized to light touches to the head by the first experimenter. The infant's eye-tracker placement occurred in one movement while the infant held the interactive toy with both hands, and the scene camera was adjusted to ensure that the toy being manipulated by the infant was in the center of the scene camera. Fifteen calibration points for the infant's eye-tracker were obtained by directing the infant's eyes toward an attractive toy and a laser pointer on the table. After placing the parent's head gear, the experimenter asked the parent to look at the laser pointing on the table in various locations to obtain fifteen calibration points for the parent's eye-tracker. During this visit, parents were instructed to play with their infant with 3 toys at a time as they would naturally do at home. Parent-infant dyads played in four 1.5-minutes-long trials, using two different sets of 3 toys in an alternating fashion across the 4 trials. The duration of the trials was chosen so that infants remained engaged in play with limited off-task behavior throughout the experiment.

The procedure at the 24-months visit consisted of an object exploration period in which children were given the toys to freely explore without any specific instructions. The experimenter sitting in front of the child limited her interactions to giving the entire set of toys to the child at the beginning of the exploration period. Parents were instructed not to engage with the objects or child. The exploration period began when the experimenter put the toys on the table and ended when the child showed no more interest in the toys on the table. Thus, the exploration period was controlled by the child.

Coding and Defined Measures

Child Gaze.

16 month visit. For each frame, the eye-tracker system generated an image with a crosshair superimposed indicating gaze at that moment. Coders used these images to determine when gaze fell in each of four regions of interest (ROIs): the three objects in play at any time and the partner's face. Each ROI was strictly defined in terms of the in-view pixels belonging to the target. The coders coded gaze –frame by frame. ROI coding was done by highly trained human coders who were naïve with respect to the research questions and who have coded these variables for multiple experiments and projects using the same dual-eye-tracking paradigm. Inter-coder reliability was high and estimated to be the same as the reliability reported by previous published research using the same coding protocols (see Yu & Smith, 2012; 2013; 2017a; 2017b).

24 month visit. Children's gaze, captured by a camera directed at the children's face was also coded frame by frame using Datavyu. The possible ROIs were the four objects on the table, the experimenter's face and the table itself. Reliability of the coder was assessed by having an independent second coder code 6 exploration periods. Both codings were compared yielding a mean Kappa score of 0.77 (SD=0.06, MIN=0.67, MAX=0.84), a scored considered to be high reliability.

Looks were defined as an *unbroken* stream of frames in which gaze was judged to fall on the same object. Duration was measured in terms of the time from the onset to the offset of this unbroken stream. In order to define episodes when parent behavior overlapped with infant unbroken looks to an object, we imposed a floor of 500 msec on the shortest infant look analyzed. In this way, the infant coded looks lasted long enough for overlapping parent behavior to occur. For the 24 month olds, who played alone with objects, we imposed no floor. For both 16- and 24-month olds, the principle measures concern **Long looks** that lasted 3 seconds or more. This threshold was defined in terms of the extremely skewed distribution of looking durations observed in the present data and is consistent with the 3-sec thresholds used for sustained attention in the literature (Ruff & Lawson, 1990; Wass et al., 2018a; Yu & Smith, 2016).

Parent behaviors- 16-months visit

Looks. Parent looks were defined as an *unbroken* stream of frames where parent gaze was judged to fall on the same object.

Talk. The speech generated by parents during the toy play at the 16-months visit was fully transcribed and divided into utterances. Each utterance was defined as a string of speech that contained vowel sounds and occurred between two periods of silence lasting at least 400 milliseconds (Suanda, Smith & Yu, 2016; Pereira, Smith & Yu, 2014; Yu & Smith, 2012). The criterion used to code the parent’s talk excluded sounds that were coughs, raspberries or sighs produced by the parent. Parents' talk included labelling utterances –such as "it's a big tower"– but also utterances that are filler sounds such as "yeah" or "uh oh." Parent talk was categorized as being about the object to which parent gaze was directed during the utterance.

Hand contact. Parent manual contact with an object at the 16-month visit was coded from images captured by the overhead camera and the other two third-person cameras. Coders accessed the three views simultaneously although in practice, coders most often relied on the overhead camera. Coders made objective frame-by-frame yes/no decisions that a parent hand was in contact with an object at that frame and did so for both hands. Inter-coder reliability was high and estimated to be the same as the reliability reported by previous published research using the same coding protocols (see Yu & Smith, 2012; 2013; 2017a; 2017b).

Parent Engagement and Scaffolded Looks. Parent engagement was determined in terms of the parent behaviors that overlapped an infant look *in time* and were directed to that same object. A parent behavior that overlapped in time with the infant's look could begin before or after the infant, and it could also be as brief as 33.3 msec , since overlap was defined when a parent behavior directed to the infant-attended object overlapped the infant look for at least 1 frame during the infant's look. This definition of overlap for parent behaviors is consistent with that used in previous research (Chapter 3).

Scaffolded looks were defined as infant looks overlapping with *High Parent Engagement* – when the parent a look, talk, and touch of the object overlapped with the infant look. Looks with *Low Parent Engagement* were looks that overlapped with some but not all three parent behaviors (e.g., only a parent look, or only a look plus a touch). Finally, looks with *No Parent Engagement* were infant looks with no overlapping parent behaviors. For some analyses, infants' looks with Low and No-parent engagement were combined into the category of non-Scaffolded looks.

Statistical Analyses

Traditional linear models were used to examine the effect of parental engagement in guiding infant attention at 16-months. Finally, correlational analyses were used to examine the relation between individual differences at 16-months in amount of long looks with High Parent Engagement and individual differences at 24-months in amount of long looks. All analyses were implemented in the R environment (version 3.3.2) (R Development Core Team, 2016).

Results

16-months Visit

Table 1 provides summary statistics of number, frequency, duration of infant looks at the corpus level and at the dyad level with looks partitioned into Short and Long and as to whether they overlapped High, Low and No Parent Engagement. Figure 2 shows the histogram of all looks in the corpus as a function of the three categories of Parent Engagement. Most infant looks were very brief (Figure 2) but on average, 20% (SD= 9%) of all infant looks passed the 3 sec threshold defining Long Looks. As expected, these Long Looks occurred most frequently in the context of High Parent Engagement (0.60 of all long looks calculated at the subject level, SD =0.20). The proportion of all looks that were long and co-occurring with High Parent Engagement (0.12, SD=0.06) was reliably greater than the proportion of all looks that were Long and co-occurring with Low Parent Engagement (0.08, SD=0.06), $t(27)=2.26$, $p=0.03$, and were also were reliably greater than the proportion of all looks that were long and co-occurred with No parent Engagement (0.00, SD=0.00), t(27)=9.47, $p<0.0001$. These results replicate previous studies (Chapter 3 as well as Wass et al., 2018a; Yu & Smith, 2016) showing that for infants, sustained visual attention is much more likely when parents look at, touch and talk about the infant-attended object.

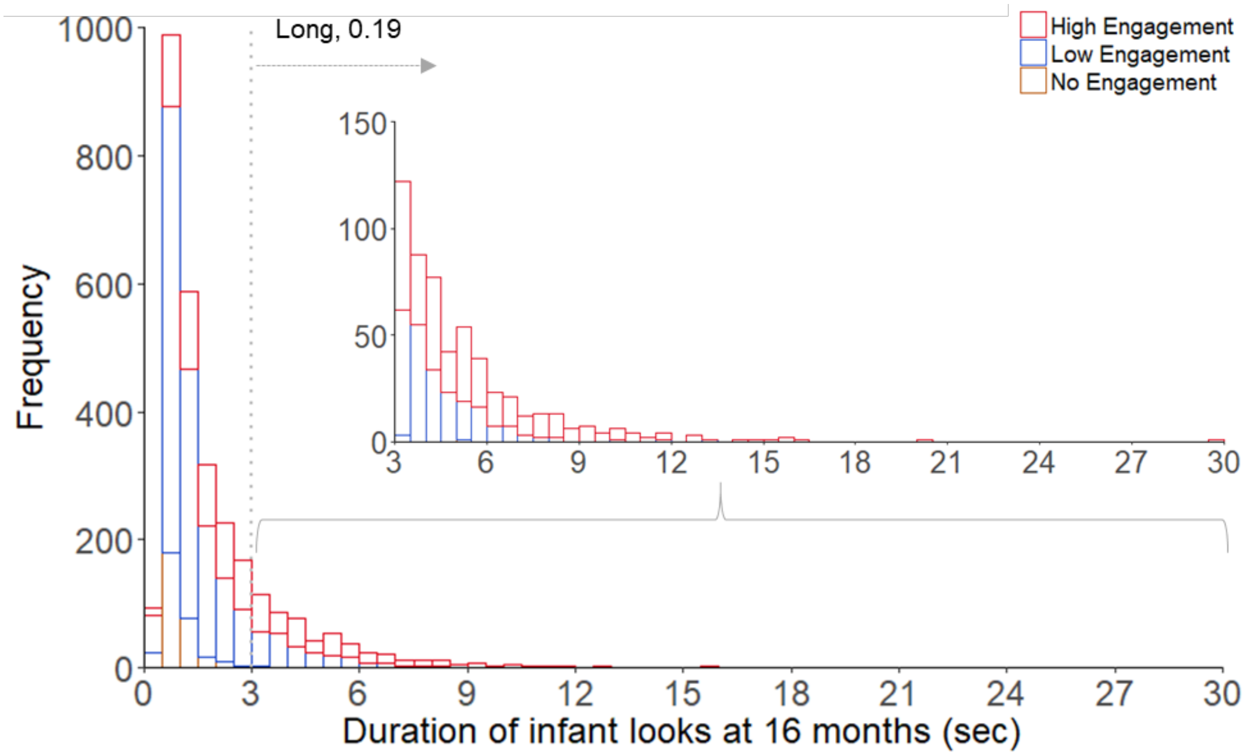


Figure 2. Frequency distribution of infant look durations at 16 months. Inset shows the durations of 19% of the data, corresponding the durations of infant looks that are Long.

Table 1.

	High Engagement	Low Engagement	No Engagement
Corpus			
Total count	818	1789	317
Count of Long	316	229	4
Count of Short	502	1560	313
Prop. Long	316/818= .39	229/1789=.13	4/317=.02

Per subject (means and descriptives of means)

Mean duration	3.36 (SD=1.02) [2.03-5.93]	1.64 (SD=0.38) [1.11-2.57]	0.96 (SD=0.21) [0.53-1.54]
Frequency (per min)	4.75 (SD=2.05) [1.49-9.53]	10.43 (SD=3.50) [4.17-16.64]	1.82 (SD=1.26) [0.15-4.60]
Prop. of all looks	0.29 (SD=0.13) [0.11-0.49]	0.61 (SD=0.09) [0.46-0.75]	0.10 (SD=0.06) [0.01-0.22]
Prop. Long looks (out of all looks)	0.12 (SD=0.06) [0.04-0.28]	0.08 (SD=0.06) [0.00-0.23]	0.00 (SD=0.00) [0.00-0.02]

As shown in Table 1 and Figure 2, there were marked individual differences in the proportions of infant looks that were Long, in the frequency of High Parent Engagement, and in the frequency of Long Looks that co-occurred with High Parent Engagement. One possibility is that High Parent Engagement in and of itself supports sustained attention. That is, an engaged social partner throughout the play session may influence infant motivation and general interest, leading to many more long looks throughout the session by the infant (Wass et al., 2018a; 2018b). A second possibility is that it is the overlap of High Parent Engagement that extends infant looks making them long (Yu & Smith, 2016). The third possibility is that infant's own tendency to sustain attention (without high parent engagement) is the key predictor of the frequency of Long Looks with High Parent Engagement. Figure 3 shows a Venn diagram that represents these three predictors of sustained attention.

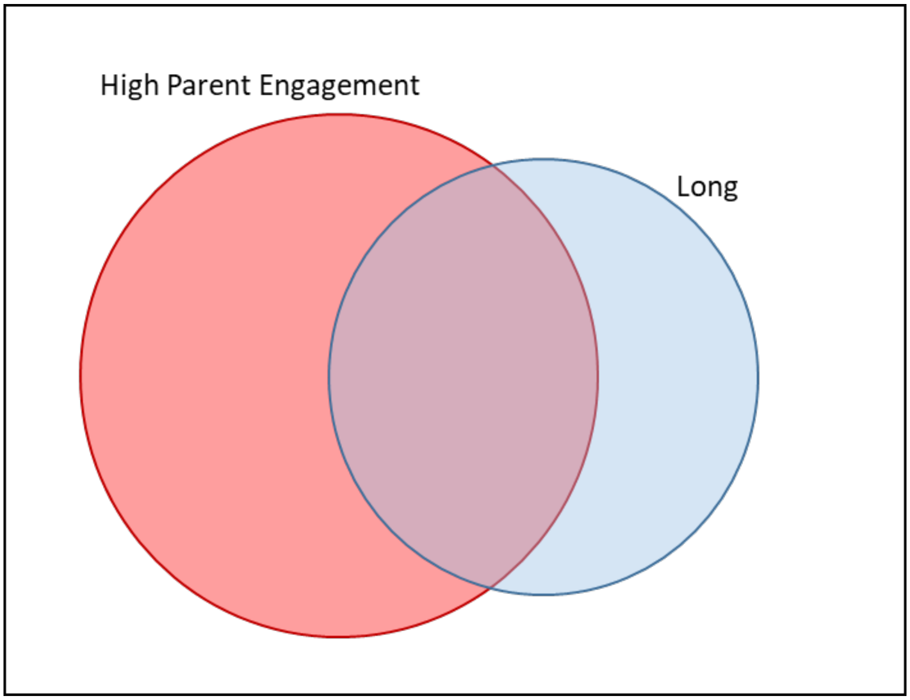


Figure 3. Number of High Parent Engagement looks, Long looks, and High Parent Engagement looks that were long represented in a Venn diagram.

The evidence at 16 months suggests that it is the overlap of High Parent Engagement with an infant look that extends the duration of that look to longer than 3 seconds. The correlation between the number of High Parent Engagement events (independent of infant looks) and the total number of an infant's Long Looks (independent of degree of parent engagement) was not reliable ($r_{pearson}$=0.20, p=0.30; See Figure 4). Further, the number of Long Looks by individual infants with and without High Parent Engagement were not correlated at 16 months, $r_{spearman}$ =0.17, p=0.39; See Figure 5).

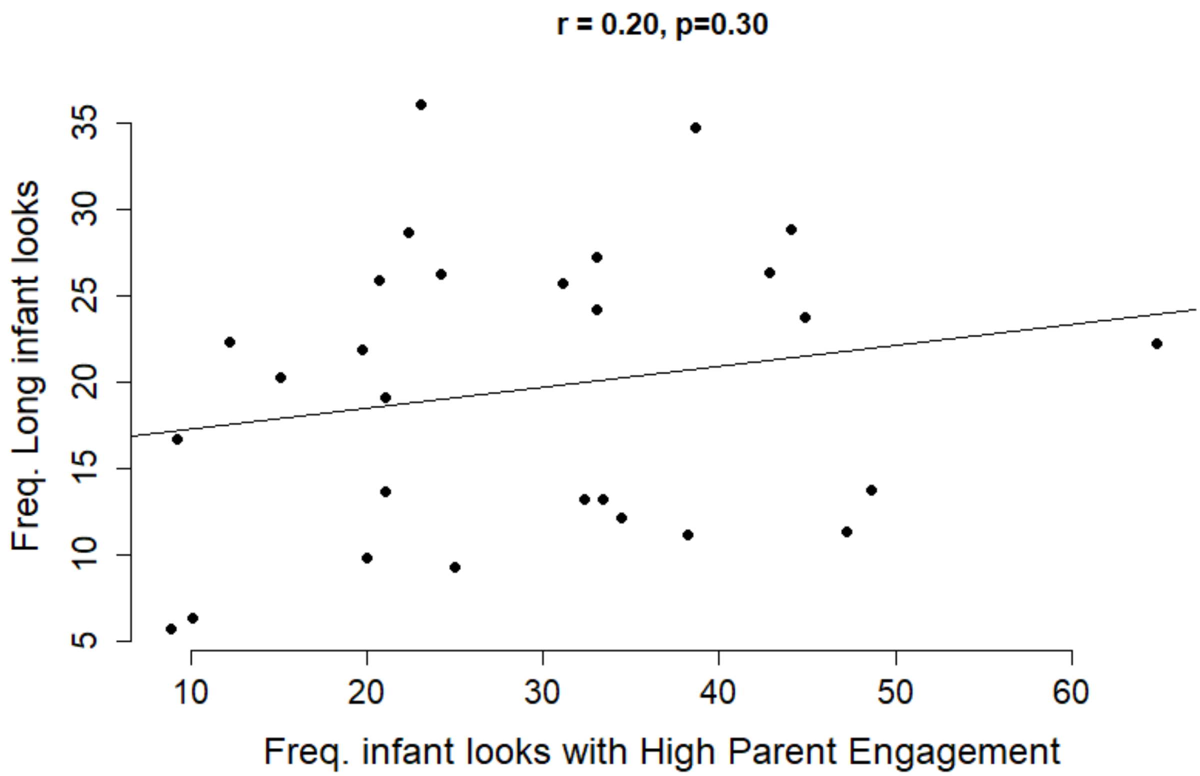


Figure 4. Frequency of Long infant looks for individual infants as a function of each infant's frequency of infant looks with High Parent Engagement.

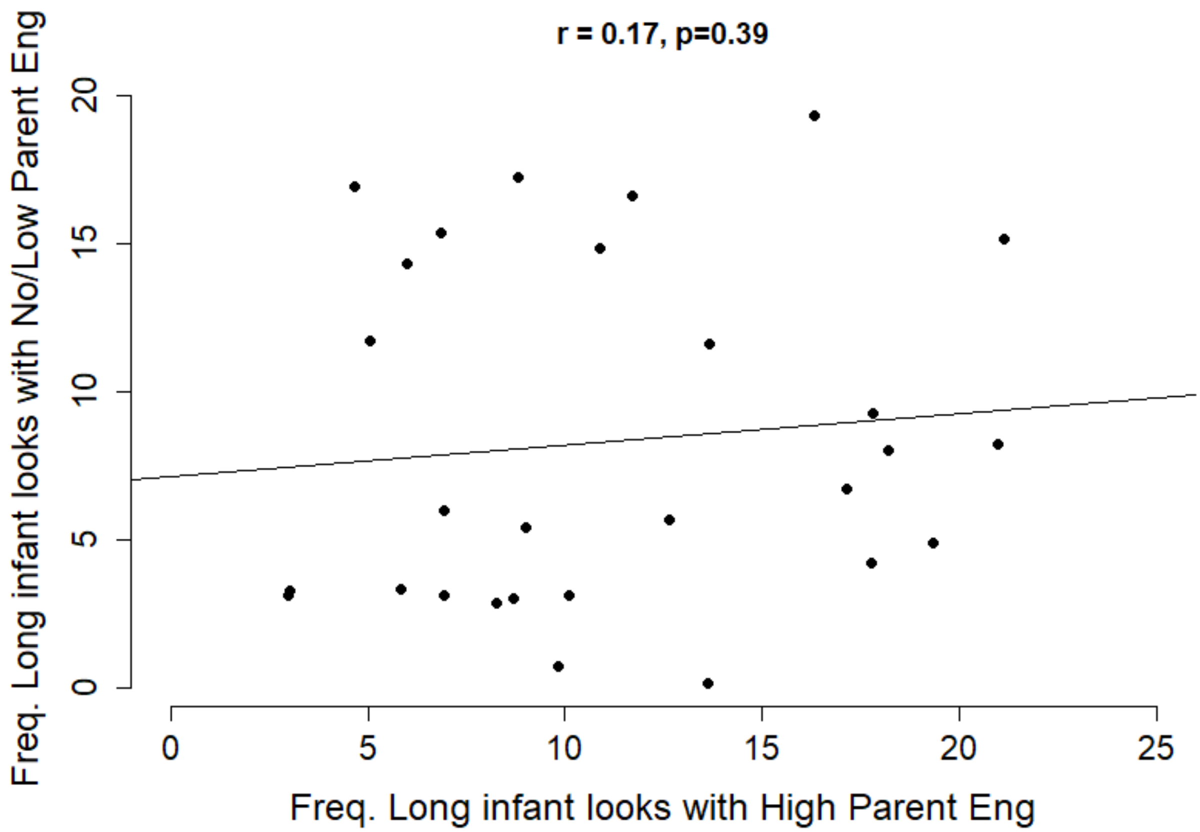


Figure 5. Frequency of Long infant looks that were Non-Scaffolded for individual infants as a function of each infant's frequency of Long infant looks that were Scaffolded.

24-months Visit

At the 24-month visit, most children visually explored all four objects that were given to them (M= 3.93, SD=0.26). The distribution of the children's look durations to individual objects is shown in Figure 6A. Again, most looks were very short, lasting less than 1 second but 19% were Long, that is longer than the 3 second threshold for sustained attention (Ruff & Lawson, 1990; Yu & Smith, 2016; Wass et al., 2018a). Figure 6B shows the count of children with different proportions of all looks that were Long looks, ranging across children from a minimum of 0.00 to maximum to 0.45 of all looks.

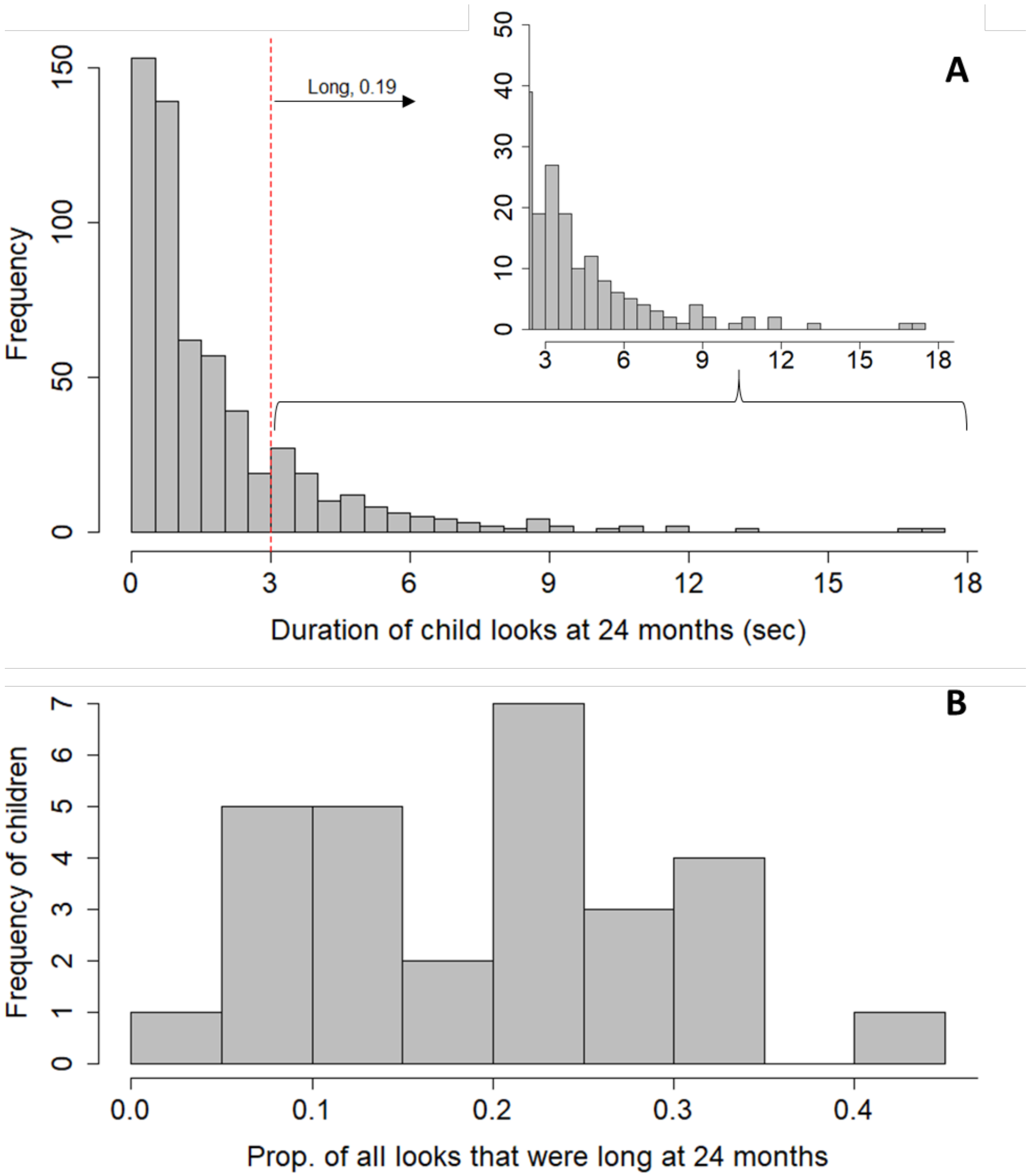


Figure 6. A. Frequency distribution of child look durations at 24 months. Inset shows the durations of 19% of the data, corresponding the durations of child attention that is Long. B. Individual differences in child Long attention at 24 months.

Consistent with the hypothesis that the early social scaffolding of long looks during parent-infant play support the development of self-controlled sustained attention, the individual differences in self-guided object exploration at 24 months were well predicted by the overlap of Social scaffolding (High parent engagement) with Long Looks at 16 months but neither by Social scaffolding alone nor by Long looks without parental scaffolding. For the following statistical analyses we used frequencies of the relevant behaviors (Scaffolded and Long, Long without scaffolding, Scaffolded regardless whether long) because the interdependencies between these measures were not reliable in terms of frequency (but were in terms of proportions). More specifically, the frequency of all looks that were long and co-occurred with High Parent Engagement at 16 months and the frequency of all looks to objects that were Long during solo play at 24 months were strongly correlated across individuals, $r_{pearson}=0.40$ $p=0.03$ (See Figure 7). The frequency of Long looks at 16 months that did not overlap with High Parent Engagement did not predict the frequency of Long looks during object exploration at 24 months ($r_{spearman} = 0.01$, $p = 0.94$) nor did the frequency of High Parent Engagement events at 16 months (independent of infant looks), $r_{pearson} = 0.17$, $p = 0.39$. In sum, the evidence supports a role for early socially-guided attention in pathway to the development of the self-regulation of visual attention: (1) 16 month old infants sustain attention to an object only in the context of parent support and (2) parents who more effectively provide that support at 16 months have older children who are more likely to sustain attention to an object on their own. Of course the present findings are correlational and thus do not unambiguously implicate causation.

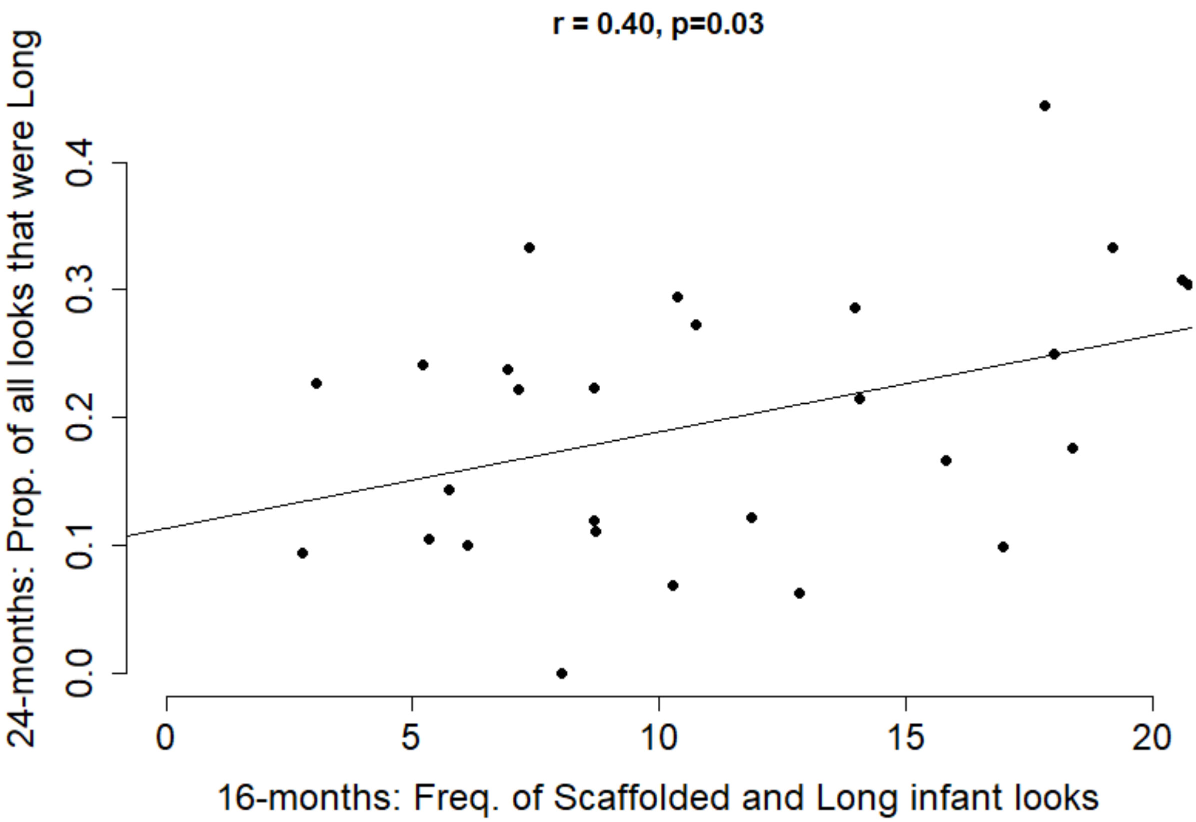


Figure 7. Individual children's abilities to sustain attention at 24 months as a function of the infant's experience with effective High Parent Engagement at 16 months.

Discussion

Young infants show more enduring visual attention to an object when parents encourage engagement with an object through their own behaviors. The present study shows that infants who experience more of these moments of socially supported infant sustained attention during play at 16 months, are those with better self-regulated attention eight months later. If early socially guided sustained attention causally supports later developing self-control of visual attention, what is the mechanism? There are at least two possible causal pathways through which moments of scaffolded attention provide the training ground that supports future attentional control.

First, moments of socially scaffolded attention may exert both exogenous and endogenous influences that extend infant attention in-the-moment and over time, provide the training ground for the self-control of attention. While Wass et al., (2018a; 2018b) showed that moments of scaffolded sustained attention occurred as a product of exogenous attentional capture and was associated with less neural activation in the infant brain, attention may result from the interaction of combined exogenous and endogenous processes (see evidence in adults presented by Ahrens, Veniero, Freund, Harvey, & Thut, 2019), and there is enough evidence to support this dual capture of attention in the context of *natural* parent-infant play. Yu & Smith (2016) showed that in the context of natural parent-infant play, overlapping Coordinated JA that extended the infant look continued to extend even after the episode of coordinated visual attention terminated, suggesting infants engaged internally controlled processes to continue the engagement. Moreover, Wass et al., (2018a) shows that individual differences in independent play are correlated with individual differences in joint play and thus showing there are endogenous drivers of attention across contexts. Finally, Wass et al., (2018a; 2018b) findings that suggest exogenous capture come from a research context that was less natural because it was a silent interaction. Findings reported in this dissertation Chapter 3 show not only that parent talk occurs in 70% of all moments of Coordinated Joint Attention but also that parent talk is the most potent driver of infant sustained attention in the context of natural play. Words are powerful drivers of endogenous attention because they activate mental representations and thus constitute a parent behavior that exerts an endogenous influence on infant attention. Accordingly, day-in day-out parents' behavior during play may engage stimuli-driven attention but it may also activates the internal systems of the infant that incrementally discover the means to control attention on their own.

Another possibility is that parent engagement with the infant-attended object activates reward systems that regulate infant behavior (Insel, 2003; Montague, Hyman & Cohen, 2004). Moments of shared engagement between the parent and the infant, the moments that extend infant attention, are rewarding because they allow infants and parents to establish a common topic of interaction and they can build a joint action around it. In the absence of joint engagement, infants may lose interest in the object, get pulled away from the object by distractors and terminate an ongoing look to move to other available objects. The reward pathway that could support the future self-control of attention may be analogous to the effect of jointly reading on independent reading later on. According to a reading analogy, reading books together does not teach children how to map sounds to written symbols but coming together around a book sets up the pattern so the behavior is more likely to occur by the child, one who is motivated to learn to read by himself. Accordingly, scaffolded sustained attention is rewarding in-the-moment and overtime it prepares system to engage with object for longer periods of time in future.

We are social animals and development happens in a social context. Research is needed to examine the exact pathways that link social partners to infant development but regardless of the pathway, the findings across multiple studies converge to suggest there is a role for social partners in nonsocial core developments. Specifically, the existing evidence consistently shows there are associations between Coordinated Joint Attention and infant sustained attention, a domain in which individual differences signal important near and long-term consequences. The study of the natural context and the specific behaviors that support this social effect is necessary for the next steps. Future research is needed to detail pathways and to carry out training studies that can make causal inferences between socially scaffolded attention and the infant's attentional control.

Supplemental Material

Figure 8 shows the frequency distribution of the latencies between the start of the infant look, and the start of the three parent behaviors –first parent look, first parent utterance, first parent touch– that overlapped in time with each infant look. *All* infant looks –both Long and Short– that co-occurred with High Parent Engagement contributed data to Figure 8. The latencies to a) the first parent look, to b) the first utterance and to c) the first touch from each infant look were calculated by subtracting the onset of the infant look a) from the onset of the parent look, b) from the onset of the parent utterance and c) from the onset of the first parent touch, respectively. As is evident in Figure 8, the latencies between infant look and all three types of parent behavior are centered around 0, and 83% of them occurred within ± 2 seconds from/to the onset of the infant's look, suggesting there is a tight and two-way coordination of infant and parent behavior. The mean latency between an infant look and parent look was 0.10 sec (SD= 0.41, median=0.11), between infant look and parent talk it was 0.07 sec (SD=0.52, median=0.13) and between infant look and parent touch it was -1.61 sec (SD=1.29, median=-1.49). These data preclude one alternative explanation of why parent behavior extends infant looks, which is that the infant attending to an object for a long period arises solely because the long duration provides an opportunity for parent behaviors to overlap the infant look. Previous studies (Suarez-Rivera et al., 2019; Wass et al., 2018a; Yu & Smith, 2016) also provide evidence against this alternative explanation of the association between sustained attention in infants and parent behaviors.

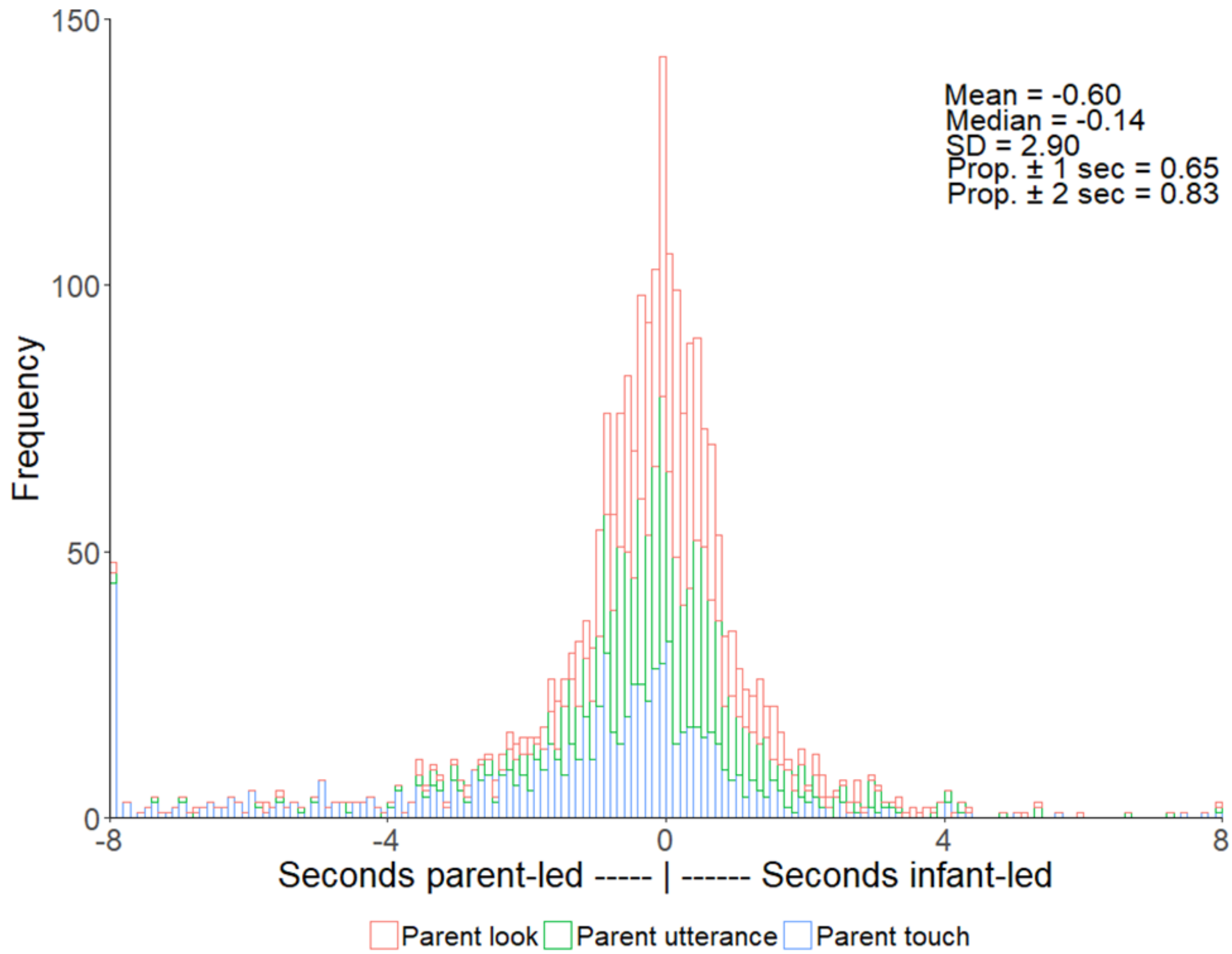


Figure 8. Latencies between infant looks with High Parent Engagement and ***each*** of the parent behaviors that temporally overlapped the infant's look (parent look (blue), parent utterance (green) and parent touch (red)). Negative latencies indicate that the parent behavior led the infant's look while positive latencies indicate the parent behavior followed the infant's look.

Chapter 5

General Discussion

This dissertation makes theoretical contributions to the study of joint attention, sustained attention and the practice of science. The studies included here not only used state of the art eye-tracking technology and microanalysis of behavior from moment-to-moment, but they also combined cross-sectional as well as longitudinal study designs in both experimental and observational research contexts. I presented strong evidence that shows the behavior elicited by trial-based gaze-following experiments is not related to a wide range of infant behaviors that occurred in the context of free flowing parent-infant play. I hypothesized about the determinants of behavior across contexts and suggested that the behavior elicited in trial-based tasks is highly dependent on the context and less representative of infant motor, cognitive, social, emotional and motivational processes because of the arbitrary decisions that researchers make when they design their tasks. The findings in Chapter 2 support the Joint Attention Disconnect hypothesis, which is just another iteration of a general problem in science. Researchers often fail to generalize findings in the lab to the complexity of the real world and the disconnect cascades into the proposal of not valid mechanisms, theories and interventions. I argued in this thesis that researchers need to move away from clean and controlled experiments because complexity and the interaction of multiple factors is part of what we need to understand in order to study development.

The last two experimental chapters focused on the more naturalistic context of parent-infant free play to examine the consequences of Coordinated Joint Attention both in real time and in developmental time for infant sustained attention. Chapter 3 reported that in natural contexts of play, Coordinated Joint Attention is a multimodal achievement in which parents not only use their eyes to engage with an infant-attended object. Instead, parents accompany these moments of Coordinated JA with other parent behaviors that signal parent interest to the infant including parent manual contact with the attended object and parent speech. This finding also shows that the phenomenon that occurs in discrete trial-based experiments is not ecologically valid as the experimenter that displays cues does not touch the object or talk about it. Moreover, the results reported in Chapter 3 showed that Coordinated JA moments in the context of natural play are strongly associated with enduring infant attention to an object, a finding that is consistent with previous reports (Wass et al., 2018a; Yu & Smith, 2016). The intriguing hypothesis is that moments of Coordinated Joint Attention that extend infant attention in the moment train the infant's system to self-control attention. I began to test this hypothesis with the study in Chapter 4. I conducted a longitudinal study with two visits one at 16 months and one at 24 months. The results showed that self-generated sustained attention at 24 months was more likely for infants that had a parent that effectively extended attention at 16 months through the constant and effective joint engagement with the infant-attended object with eyes, hands and talk. Infants who sustained attention without the parent support more and infants who had constant engagement with the parent overall, were infants who did not engaged in a lot of self-generated sustained attention at 24 months. I proposed plausible pathways that can directly link socially scaffolding of attention with the future self-control of attention. I discussed the existing literature on the internal endogenous vs. exogenous mechanisms that support socially guided sustained attention, in light of the research context that generated the findings. I suggested that it is possible that infant sustained attention that occurs with a parent during play may be captured by both endogenous and exogenous mechanisms. The study implicates social partners in the development of the self-regulation of attention but more work is needed to make causal inferences about Coordinate Joint Attention and the drivers of attention.

The studies presented in the dissertation are connected because together they build a pathway from Coordinated JA, to Sustained Attention and to Later Outcomes. This pathway is one that could detail the specific ways in which social partners matter for the development of infants because previous research has already shown that social partners matter for development (Bornstein & Tamis-Lemonda, 1997). There are important real-world applications of this pathway. The evidence can inform and change the interventions that are currently being used to improve gaze-following in children with ASD (Carlson, Wong, Dung, Wong, Tan & Wykowska, 2018; Miller, Wyatt, Casey & Smith, 2018). In addition, the research included in this dissertation has the potential to inform interventions for parents of typically developing infants. If parents extend infant attention in-the-moment, then they may support the development of the child's own control of attention. It is possible that by training parental support of infant attention, children may not only gain a better control of attention but better language abilities and self-regulation skills.

The evidence presented on the socially supported sustained attention and the development of self-generated attention, is purely correlational and causal inferences cannot be made at this point. Another limitation for all studies is that the context of free play used took place in the laboratory and it is an open question whether it is representative of all aspects of the interactions that take place in the home environment. In the same way, the participants were mostly middle-class white families and generalization to the entire population is not certain. Finally, the test-retest reliability of measures obtained in the free play context is unknown but they are needed.

Future work is needed to examine parent-infant joint engagement in the context of the home, in more real-world contexts that could allow researchers to estimate how frequent these moments are and whether the proposed pathways that link Coordinated JA to the development of self-generated sustained attention may be plausible. In the context of the home, parents have competing tasks and daily routines as opposed to the context of play in the lab in which parents are only focused on the play session and infant's attentional state. How often do moments of Coordinated Joint Attention occur in the home and what are the behaviors used in the home context to support sustained attention? In the same way, future work should carry out a training study to test the causal relation between socially scaffolded attention and the development of self-generated attention. In a training study, infants would be tested in the context of independent play with object both before and after the training session, which would consist of between 3-8 sessions with an experimenter who is trained to support attention by generating multimodal behaviors around the infant-attended object. The gains in the ability to sustain attention for this group of infants can be compared to the gains obtained by infants in a control group who came to the lab but did not engage in the social scaffolding of sustained attention. This type of study could still elicit natural behavior but causally test whether social partners can promote better self-generated attention. It could also speak to how malleable the path is, an open question that needs to be answered before interventions are implemented.

To conclude, I pose some of the open questions for the field suggested by this thesis. What are the gaze-following discrete trial-based experiments measuring? The results showed the discrete-trial based experiment had large individual differences and infants followed the gaze cue very quickly. Because these measured individual differences have been shown to be predictive of future development, it is important to ask whether the variance captured relates to any other tasks used in the field to measure other constructs such as disengagement, executive function or recovery in the face of perturbations.

Another important question is whether researchers are also unaware of other trial-based tasks in the literature that also elicit a non-natural behavior in the context of the experiment. Developmental psychology is dominated by experimental studies that isolate factors and try to find the single cause of learning and development. These “petri dish” studies are likely missing the complexities between more than one factor that in real life interact and matter for learning and development. It is important that researchers stop calling mechanisms to the surface behaviors that are shown in lab-based highly predictive tasks of future development.

Finally, research about research is needed to understand how to carry out research that retains ecological validity and elicits real-world phenomena, but that also helps us understand mechanisms of action. Even if it is shown that experiments are not bulletproof and that causal inference should not be the only goal of research (Chapter 2), researchers and the general public will always want to understand how things work (Chapter 3 and 4). Is there a gold standard model that indicates how to combine observational and experimental research designs? Perhaps the optimal combination of methods could accomplish the right balance between ecological validity and causal inferences. Moreover, the research enterprise as a whole places an emphasis on the reliability of the experimental work by asking researchers to report reliability of the measures used. But, should we also care about the validity of the research? In particular, should researchers be asked to explicitly examine or discuss the ecological validity of their experiments? Answers to these questions could help current and future generations of scientists make important contributions to our understanding of the world through the implementation of strong research methods.

References

Ahrens, M. M., Veniero, D., Freund, I. M., Harvey, M., & Thut, G. (2019). Both Dorsal and Ventral Attention Network Nodes are implicated in Exogenously Driven Visuospatial Anticipation. *Cortex*.

Akhtar, N., & Gernsbacher, M. A. (2007). Joint attention and vocabulary development: A critical look. *Language and linguistics compass*, *1*(3), 195-207.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 1278-1289. https://doi.org/10.2307/1129997

Baldwin, D. A., & Markman, E. M. (1989). Establishing word-object relations: A first step. *Child Development*, 381-398. https://doi.org/10.2307/1130984

Baron-Cohen, S., & Cross, P. (1992). Reading the eyes: evidence for the role of perception in the development of a theory of mind. *Mind & Language, 7*(1□2), 172-186. https://doi.org/10.1111/j.1468-0017.1992.tb00203.x

Bates, D., Maechler, M., Bolker,B., Walker. S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J*ournal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01. https://doi.org/10.18637/jss.v067.i01

Belsky, J., Goode, M. K., & Most, R. K. (1980). Maternal stimulation and infant exploratory competence: Cross-sectional, correlational, and experimental analyses. *Child Development*, 1168-1178. https://doi.org/10.2307/1129558

Bornstein, M. H., & Tamis-Lemonda, C. S. (1997). Maternal responsiveness and infant mental abilities: Specific predictive relations. *Infant Behavior and Development, 20*(3), 283-296. https://doi.org/10.1016/S0163-6383(97)90001-1

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science, 8*(6), 535-543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, *33*(2), 261-304.

Carpenter, R. H. (1988). *Movements of the Eyes, 2nd Rev*. Pion Limited.

Carlson, K., Wong, A. H. Y., Dung, T. A., Wong, A. C. Y., Tan, Y. K., & Wykowska, A. (2018, November). Training Autistic Children on Joint Attention Skills with a Robot. In *International Conference on Social Robotics* (pp. 86-92). Springer, Cham.

Carvalho, P. F., Vales, C., Fausey, C. M., & Smith, L. B. (2018). Novel names extend for how long preschool children sample visual information. J*ournal of experimental child psychology*, 168, 1-18. https://doi.org/10.1016/j.jecp.2017.12.002

Charman, T. (2003). Why is joint attention a pivotal skill in autism?. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1430), 315-324.

Clauset, A., Shalizi, C. R., & Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review, 51*(4), 661-703. https://doi.org/10.1137/070710111

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Phil. Trans. R. Soc. B, 372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

Colombo, J. (2001). The development of visual attention in infancy. *Annual review of psychology, 52*(1), 337-367. https://doi.org/10.1146/annurev.psych.52.1.337

Colombo, J., & Mitchell, D. W. (1990). Individual differences in early visual attention: Fixation time and information processing. *Individual differences in infancy: Reliability, stability, prediction*, 193-227.

Corkum V, Moore C (1998) Origins of joint visual attention in infants. *Developmental Psychology* 34: 28. https://doi.org/10.1037/0012-1649.34.1.28

Dahl, A. (2017). Ecological commitments: Why developmental science needs naturalistic methods. *Child development perspectives*, *11*(2), 79-84.

Deák, G. O. (2000). Hunting the fox of word learning: Why "constraints" fail to capture it. *Developmental Review*, *20*(1), 29-80.

Deak, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science, 17*(2), 270-281. https://doi.org/10.1111/desc.12122

Doherty MJ, Anderson JR, Howieson L (2009) The rapid development of explicit gaze judgment ability at 3 years. *Journal of experimental child psychology*, 104: 296–312. https://doi.org/10.1016/j.jecp.2009.06.004

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C., (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428. https://doi.org/10.1037/0012-1649.43.6.1428

Farroni T, Johnson MH, Brockbank M, Simion F (2000) Infants' use of gaze direction to cue attention: The importance of perceived motion. *Visual Cognition,* 7: 705–718. https://doi.org/10.1080/13506280050144399

Fausey, C. M., Jayaraman, S., & Smith, L. B. (2016). From faces to hands: Changing visual input in the first two years. *Cognition*, *152*, 101-107.

Fay-Stammbach, T., Hawes, D. J., & Meredith, P. (2014). Parenting influences on executive function in early childhood: A review. *Child Development Perspectives*, *8*(4), 258-264.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology*, *42*(1), 98.

Fernald, A., Thorpe, K., & Marchman, V. A. (2010). Blue car, red car: Developing efficiency in online interpretation of adjective–noun phrases. *Cognitive Psychology, 60*(3), 190-217. https://doi.org/10.1016/j.cogpsych.2009.12.002

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research, 50*(24), 2766-2774. https://doi.org/10.1016/j.visres.2010.09.024

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head‐mounted eye tracking: A new method to describe infant looking. *Child Development, 82*(6), 1738-1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: Do observational findings reflect the natural behavior of participants?. *Clinical child and family psychology review*, *3*(3), 185-198.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274-279. https://doi.org/10.1111/1467-9280.00255

Grolnick, W. S., & Ryan, R. M. (1989). Parent styles associated with children's self-regulation and competence in school. *Journal of educational psychology*, *81*(2), 143.

Gros-Louis, J., West, M. J., Gold, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, *30*(6), 509-516.

Henderson, J. M., & Ferreira, F. (2004). Scene Perception for Psycholinguists.

Insel, T. R. (2003). Is social attachment an addictive disorder?. *Physiology & Behavior, 79*(3), 351-357. https://doi.org/10.1016/S0031-9384(03)00148-3

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2017). Why are faces denser in the visual experiences of younger than older infants?. *Developmental psychology*, *53*(1), 38.

Johnson, E. K., McQueen, J. M., & Huettig, F. (2011). Toddlers' language-mediated visual search: They need not have the words for it. *Quarterly Journal of Experimental Psychology, 64*(9), 1672-1682. https://doi.org/10.1080/17470218.2011.594165

Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E., & Bornstein, M. H. (2015). Places and postures: A cross-cultural comparison of sitting in 5-month-olds. *Journal of cross-cultural psychology*, *46*(8), 1023-1038.

Kello, C. T., Brown, G. D., Ferrer-i-Cancho, R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., & Van Orden, G. C. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences, 14(*5), 223-232. https://doi.org/10.1016/j.tics.2010.02.005

Kessler, R., & Glasgow, R. E. (2011). A proposal to speed translation of healthcare research into practice: dramatic change is needed. *American journal of preventive medicine*, *40*(6), 637-644.

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*(2), 220. https://doi.org/10.1037/0012-1649.36.2.220

Kopp, C. B. (1982). Antecedents of self-regulation: a developmental perspective. *Developmental Psychology, 18*(2), 199. https://doi.org/10.1037/0012-1649.18.2.199

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives*, *9*(2), 74-78.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities?. *Vision research, 41*(25-26), 3559-3565. https://doi.org/10.1016/S0042-6989(01)00102-X

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*(11), 1311-1328. https://doi.org/10.1068/p2935.

Landau, B., Smith, L. B., & Jones, S. (1992). Syntactic context and the shape bias in children's and adults' lexical learning. J*ournal of Memory and Language, 31*(6), 807-825. https://doi.org/10.1016/0749-596X(92)90040-5

Landry, S. H., Smith, K. E., & Swank, P. R. (2006). Responsive parenting: establishing early foundations for social, communication, and independent problem-solving skills. *Developmental psychology*, *42*(4), 627.

Langton SRH, Watt RJ, Bruce V (2000) Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences,* 4: 50–59. https://doi.org/10.1016/S1364-6613(99)01436-9

Lansink, J. M., & Richards, J. E. (1997). Heart Rate and Behavioral measures of Attention in Six□, Nine, and Twelve□Month□Old Infants during Object Exploration. *Child Development, 68*(4), 610-620. https://doi.org/10.2307/1132113

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software, 69*(1), 1-33. https://doi.org/10.18637/jss.v069.i01

Loomis JM, Kelly JW, Pusch M, Bailenson JN, Beall AC (2008) Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, 37: 1443–1457. https://doi.org/10.1068/p5896

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders*, *30*(3), 205-223.

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive psychology*, *47*(3), 241-275.

Markus, J., Mundy, P., Morales, M., Delgado, C. E., & Yale, M. (2000). Individual differences in infant skills as predictors of child□caregiver joint attention and language. *Social Development*, *9*(3), 302-315.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*(2), 320-339. https://doi.org/10.1111/j.1756-8765.2009.01022.x

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*(1), 60-76.

Mather, E., & Plunkett, K. (2011). Same items, different order: Effects of temporal variability on infant categorization. *Cognition*, *119*(3), 438-447.

McClelland, M. M., & Cameron, C. E. (2012). Self□regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives, 6(*2), 136-142. https://doi.org/10.1111/j.1750-8606.2011.00191.x

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21*(4), 471-490. https://doi.org/10.1016/j.ecresq.2006.09.003

Miller, J. L., Ables, E. M., King, A. P., & West, M. J. (2009). Different patterns of contingent stimulation differentially affect attention span in prelinguistic infants. *Infant Behavior and Development, 32*(3), 254-261. https://doi.org/10.1016/j.infbeh.2009.02.003.

Miller, N., Wyatt, J., Casey, L. B., & Smith, J. B. (2018). Using computer□assisted instruction to increase the eye gaze of children with autism. *Behavioral Interventions*, *33*(1), 3-12.

Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature, 431*(7010), 760. https://doi.org/10.1038/nature03015

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373-377.

Morales, M., Mundy, P., Crowson, M., Neal, A. R., & Delgado, C. (2005). Individual differences in infant attention skills, joint attention, and emotion regulation behaviour. *International Journal of Behavioral Development*, *29*(3), 259-263.

Mundy, P., Card, J., & Fox, N. (2000). EEG correlates of the development of infant joint attention skills. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *36*(4), 325-338.

Mundy, P., Delgado, C., Block, J., Venezia, M., Hogan, A., & Seibert, J. (2003). Early social communication scales (ESCS). *Coral Gables, FL: University of Miami*.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Van Hecke, A. V., & Parlade, M. V. (2007). Individual differences and the development of joint attention in infancy. *Child development, 78*(3), 938-954. https://doi.org/10.1111/j.1467-8624.2007.01042.x

Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed processing model of joint attention, social cognition and autism. *Autism research*, *2*(1), 2-21. https://doi.org/10.1002/aur.61

Oakes, L. M. (2015). A biopsychosocial perspective on looking behavior in infancy. *Handbook of infant biopsychosocial development*, 38-70.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*(4), 601-626. https://doi.org/10.1016/j.jml.2006.08.009

Parrinello, R. M., & Ruff, H. A. (1988). The influence of adult intervention on infants' level of attention. *Child development,* 1125-1135. https://doi.org/10.2307/1130279

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion*, *9*(4), 560.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review, 21*(1), 178-185. https://doi.org/10.3758/s13423-013-0466-4

Pérez-Edgar, K., McDermott, J. N. M., Korelitz, K., Degnan, K. A., Curby, T. W., Pine, D. S., & Fox, N. A. (2010). Patterns of sustained attention in infancy shape the developmental trajectory of social behavior from toddlerhood through adolescence. *Developmental Psychology, 46*(6), 1723. https://doi.org/10.1037/a0021064

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112-1130. https://doi.org/10.3758/s13423-014-0585-6

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus,* 3-56. https://doi.org/10.1007/978-1-4419-0318-1_1

Posner, M. I., & Rothbart, M. K. (2000). Developing mechanisms of self-regulation. *Development and Psychopathology, 12(*3), 427-441. https://doi.org/10.1017/S0954579400003096

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reck, S. G., & Hund, A. M. (2011). Sustained attention and age predict inhibitory control during early childhood. *Journal of experimental child psychology, 108*(3), 504-512. https://doi.org/10.1016/j.jecp.2010.07.010

Richards, J. E., & Casey, B. J. (1992). Development of sustained visual attention in the human infant. *Attention and information processing in infants and adults: Perspectives from human and animal research*, 30-60.

Riggs, B. L., Hodgson, S. F., O'fallon, W. M., Chao, E. Y., Wahner, H. W., Muhs, J. M., Cedel, S. L., & Melon III, L. J. (1990). Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *New England journal of medicine*, *322*(12), 802-809.

Rothwell, P. M. (2005). External validity of randomised controlled trials:"to whom do the results of this trial apply?". *The Lancet*, *365*(9453), 82-93.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663-12668.

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child Development*, 105-114. https://doi.org/10.2307/1130642

Ruff, H. A. (1990). Individual differences in sustained attention during infancy. *Individual Differences in Infancy: Reliability, Stability, Prediction,* 247-270.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental Psychology, 26*(1), 85. https://doi.org/10.1037/0012-1649.26.1.85

Ruff, H. A., Lawson, K. R., Parrinello, R., & Weissberg, R. (1990). Long□term stability of individual differences in sustained attention in the early years. *Child development*, *61*(1), 60-75.

Sigel, Irving E. "The psychological distancing model: A study of the socialization of cognition." *Culture & Psychology 8.2 (2002)*: 189-214. https://doi.org/10.1177/1354067X02008002438

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in cognitive sciences*, *7*(8), 343-348.

Smith, C. B., Adamson, L. B., & Bakeman, R. (1988). Interactional predictors of early language. *First Language*, *8*(23), 143-156.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419-436.

Spruijt, A. M., Dekker, M. C., Ziermans, T. B., & Swaab, H. (2018). Attentional control and executive functioning in school-aged children: Linking self-regulation and parenting strategies. *Journal of Experimental Child Psychology, 166,* 340-359. https://doi.org/10.1016/j.jecp.2017.09.004

Steinmayr, R., Ziegler, M., & Träuble, B. (2010). Do intelligence and sustained attention interact in predicting academic achievement?. *Learning and Individual Differences, 20*(1), 14-18. https://doi.org/10.1016/j.lindif.2009.10.009

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The Multisensory Nature of Verbal Discourse in Parent–Toddler Interactions. *Developmental Neuropsychology, 41*(5-8), 324-341. https://doi.org/10.1080/87565641.2016.1256403

Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology, 23*(21), 2162-2168. https://doi.org/10.1016/j.cub.2013.09.005

Tamis□LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child Development, 72*(3), 748-767. https://doi.org/10.1111/1467-8624.00313

Tamis□LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2018). Routine language: Speech directed to infants during home activities. *Child development*.

Tamis□LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental science*, *20*(6), e12456.

Tasker, S. L., & Schmidt, L. A. (2008). The “dual usage problem” in the explanations of “joint attention” and children’s socioemotional development: A reconceptualization. *Developmental Review*, *28*(3), 263-288.

Team, D. (2014). Datavyu: A video coding tool. *Databrary Project, New York University. URL http://datavyu. org*.

Teo, S. S. S., & Shingadia, D. V. (2006). Does BCG have a role in tuberculosis control and prevention in the United Kingdom?. *Archives of disease in childhood*, *91*(6), 529-531.

Tomasello, M. (1995). Joint attention as social cognition. *Joint attention: Its origins and role in development*, *103130*.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development,* 1454-1463. https://doi.org/10.2307/1130423

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, *4*(12), 197-211.

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117-135. https://doi.org/10.1016/j.cognition.2015.11.002

Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3□year□old children. *Developmental Science, 18*(1), 65-79. https://doi.org/10.1111/desc.12179

Van Egeren, L. A., Barratt, M. S., & Roach, M. A. (2001). Mother–infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology, 37*(5), 684. https://doi.org/10.1037/0012-1649.37.5.684

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., Soderstrom, M., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. In *Seminars in speech and language* (Vol. 37, No. 02, pp. 128-142). Thieme Medical Publishers.

Vaughan, A., Mundy, P., Block, J., Burnette, C., Delgado, C., Gomez, Y., Meyer, J., Neal, A. R., & Pomares, Y. (2003). Child, caregiver, and temperament contributions to infant joint attention. *Infancy*, *4*(4), 603-616.

Vida M, Maurer D (2012a). Fine-grained sensitivity to vertical differences in triadic gaze is slow to develop. *Journal of Vision*, 12: 634–634. https://doi.org/10.1167/12.9.634

Vida MD, Maurer D (2012b). The development of fine-grained sensitivity to eye contact after 6 years of age. *Journal of experimental child psychology,* 112: 243–256. https://doi.org/10.1016/j.jecp.2012.02.002

Vida MD, Maurer D (2012c). Gradual improvement in fine-grained sensitivity to triadic gaze after 6 years of age. *Journal of experimental child psychology,* 111: 299–318. https://doi.org/10.1016/j.jecp.2011.08.009

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children, 23*(3), 34-41.

Wass, S. V., Clackson, K., Georgieva, S. D., Brightman, L., Nutbrown, R., & Leong, V. (2018a). Infants' visual sustained attention is higher during joint play than solo play: is this due to increased endogenous attention control or exogenous stimulus capture?. *Developmental science*, e12667. https://doi.org/10.1111/desc.12667

Wass, S. V., Noreika, V., Georgieva, S., Clackson, K., Brightman, L., Nutbrown, R., Covarrubias, L. S., & Leong, V. (2018b). Parental neural responsivity to infants' visual attention: how mature brains influence immature brains during social interaction. *PLoS biology*, *16*(12), e2006328.

Wei, F. Y. F., Wang, Y. K., & Klausner, M. (2012). Rethinking college students' self-regulation and sustained attention: Does text messaging during class influence cognitive learning?. *Communication Education, 61*(3), 185-204. https://doi.org/10.1080/03634523.2012.672755

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology, 102(*1), 43. https://doi.org/10.1037/a0016738

Yoshida H, Smith LB (2008) What's in view for toddlers? Using a head camera to study visual experience. *Infancy* 13: 229–248. https://doi.org/10.1080/15250000802004437

Yu, C., & Smith, L.B.(2012). Embodied Attention and Word Learning by Toddlers. *Cognition, 125*(2), 244-262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one, 8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology, 26*(9), 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., & Smith, L. B. (2017a). Multiple Sensory□Motor Pathways Lead to Coordinated Visual Attention. *Cognitive Science, 41*(S1), 5-31. https://doi.org/10.1111/cogs.12366

Yu, C., & Smith, L. B. (2017b). Hand–eye coordination predicts joint attention. *Child development*, *88*(6), 2060-2078.

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental science*, *22*(1), e12735.

Catalina Suarez Rivera

Curriculum vitae

CONTACT Indiana University
Department of Psychological and Brain Sciences
1101 E. 10th St, Bloomington, IN 47405
csuarezr@indiana.edu
https://csuarezr7.wixsite.com/suarezrivera

EDUCATION

2014-2019 Ph.D. Psychology, MS in Applied Statistics, Indiana University: Bloomington, IN
Cognitive Development Lab (P.I: Linda Smith)
Computational Cognition and Learning Laboratory (P.I: Chen Yu)

2011-2013 B.S Psychology *summa cum laude*, Illinois Institute of Technology: Chicago, IL
Social Communication Network Lab (P.I: Jennifer Miller)

2010 Psychology, University of The Andes: Bogota, Colombia

RESEARCH INTERESTS

Implications of parent-infant interactions for the development of infant attention and language

AWARDS

2019 Provost's Travel Award for Women in Science- Indiana University
2017 Graduate and Professional Student Government Travel Award- Indiana University
2016 Young Scientist Travel Award- Indiana University
2016 Simons Foundation Travel Award- International Congress of Infant Studies
2015 Provost's Travel Award for Women in Science- Indiana University
2014-Present Graduate Assistantship- Indiana University
2013 Undergraduate Student Leadership Award- Illinois Institute of Technology
2013 Best Psychology Poster- Chicago Area Undergraduate Research Symposium
2011-2013 International Student Scholarship- IIT
2010-2013 Graduates for Colombia Scholarship- Ecopetrol
2011-2013 Dean's List- Department of Psychology, IIT
2012 Research Scholarship- Department of Psychology, IIT
2010 I Want to Go to College Scholarship- University of the Andes

PUBLICATIONS

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology, 55*(1), 96-109. http://dx.doi.org/10.1037/dev0000628

Slone, K. L., Abney, D. H., Borjon, J. I., Chen, C., Pearcy, D., **Suarez-Rivera, C**., Zhang, Y., Smith, L. B., Yu, C. (2018). Gaze in action: Head-mounted eye-tracking of children's dynamic visual attention during naturalistic behavior. *Journal of Visualized Experiments,* (141)*,* e58496. http://dx.doi.org/10.3791/58496.

Miller, J. L., Lossia, A., **Suarez-Rivera, C.**, & Gros-Louis, J. (2017). Toys that squeak: Toy type impacts quality and quantity of parent–child interactions. *First Language*, *37*(6), 630-647. https://doi.org/10.1177/0142723717714947

MANUSCRIPTS IN PREPARATION

Suarez-Rivera, C., Smith, L. B., & Yu, C. (in prep). *Socially scaffolded sustained attention predicts future independent sustained visual attention.*

ORAL COMMUNICATIONS

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019, March). *Long term consequences of joint parent-infant play: Social partners train infant's regulation of attention.* Communication presented at the Society for Research in Child Development Biennial Meeting, Baltimore, MD, USA.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2018, July). *From faces to objects: crawling experience modifies patterns of infant attention.* Communication presented at the International Congress of Infant Studies, Philadelphia, PA, USA.

***Suarez-Rivera, C.**, & Slone, L. K. (co-presenters, 2017, August). *Seeing things from the infants' view: How infant and caregiver engagement during play promotes early learning.* Talk presented at Bridging Early Connections: Indiana Infant Toddler Institute, Indianapolis, IN, USA.

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2016, May). *The social origins of sustained attention*. Communication presented at the International Congress of Infant Studies, New Orleans, LA, USA.

POSTER COMMUNICATIONS

Yurkovic, J. R., **Suarez-Rivera, C.,** Lisandrelli, G., Shaffer, R., Pedapati, E., Erickson, C. A., Yu, C., Kennedy, D. P. (2019, May). *The Effects of Multimodal Behaviors from Parents on Children's Sustained Attention: A Dual Eye Tracking Study during Naturalistic Child-Parent Social Interaction.* Poster presented at the International Society for Autism Research 2019 Annual Meeting, Montréal, Canada.

Suarez-Rivera, C., Smith, L. B., & Yu, C (2019, March). *The joint attention disconnect: Comparing babies in the wild and babies in the lab.* Poster presented at the Society for Research in Child Development Biennial Meeting, Baltimore, MD, USA.

Suarez-Rivera, C., Mankovich, A., Smith, L. B., & Yu, C (2017, March). *Infant sustained attention: A role of parent talk.* Poster presented at the International Convention of Psychological Science, Vienna, Austria.

Lossia, A. K., **Suarez-Rivera, C.**, & Miller, J. L. (2013, November). *Triad Census: A new method for analyzing caregiver-infant interactions.* Poster presented at Social Networks and Innovation: A Conference on New Frontiers in Methods and Application, Chicago, IL, USA.

McCue, K., Reife, I., Dorsaint, T., **Suarez-Rivera, C.**, Pichette, C., & Wang, T. (2013, May). *Anxiety symptoms as predictor of adaptive functioning in ASD.* Poster presented at the Midwestern Psychological Association Conference, Chicago, IL, USA.

Suarez-Rivera, C., Lossia, A. K., & Miller, J. L. (2013, April). *Toys that talk: The influence of toy properties on prelinguistic communicative behaviors.* Poster presented at the Chicago Area Undergraduate Research Symposium, Chicago, IL, USA.

Lossia, A. K., **Suarez-Rivera, C.** & Miller, J. L. (2013, April). *Toys that talk: The influence of toy properties on prelinguistic communicative behaviors.* Poster presented at the Society for Research in Child Development Biennial Meeting, Seattle, WA, USA.

Miller, J. L., Lossia, A. K., & **Suarez-Rivera, C.** (2013, April). *Triadic Census: A new method to analyzing complex social interactions.* Poster presented at the Society for Research in Child Development Biennial Meeting, Seattle, WA, USA.

PROFESSIONAL SERVICE

AD HOC REVIEWER

Developmental Science

TEACHING EXPERIENCE

Fall 2018 *Teaching Assistant*
Psychobiology, Self and Society (PSY 457), Indiana University

Spring 2018 *Teaching Assistant*
Developmental Psychology (PSY 315), Indiana University

Fall 2017 *Lab Instructor: ratings 3.41/4*
Statistical Techniques (PSY K300), Indiana University

Spring 2017 *Lab Instructor: ratings 3.55/4*
Introduction to Research Methods in Psychology (PSY 211), Indiana University

Spring 2016 *Teaching Assistant*
Statistical Techniques (PSY K300), Indiana University

Fall 2015 *Teaching Assistant*
Human Learning and Cognition Lab (P435), Indiana University

Fall 14-Spring 15 *Teaching Assistant*
Statistical Techniques (PSY K300), Indiana University

MENTORSHIP EXPERIENCE

2016-2018 Marissa Tabor, Indiana University

2017 Nora Pendergast, Indiana University

2017 Rebecca Wilkinson, Indiana University

2016-2017 Daniel Piercy (IU, lab manager)

2015-2016 Amanda Mankovich (University of Connecticut, PhD student)

2015-2016 Maria Gregg, Indiana University

2015 Jonathan Roskam, Indiana University

2015 Stephanie Irwin, Indiana University

2015 Cheyenne Reyes, Indiana University

PROFESSIONAL ORGANIZATIONS

International Congress of Infant Studies

Society for Research in Child Development

REFERENCES

References available upon request